

## کاربردی از مدل آمیخته‌ی خطی در برآورد کوچک ناحیه‌ای محصول پرتقال در استان فارس

حمیدرضا نواب پور<sup>۱</sup>

چکیده:

در سال‌های اخیر روش‌های برآورد کوچک ناحیه‌ای مورد توجه قرار گرفته‌اند. این توجه به دلیل افزایش درخواست برآوردهای معتبر برای کوچک ناحیه‌ها بوده است، زیرا این برآوردها برای برنامه‌ریزی‌های توسعه‌ای دقیق‌تر، تخصیص اعتبارهای دولتی و تصمیم‌گیری‌های تجاری به‌کار می‌آیند. پرسش کلیدی در برآورد کوچک ناحیه‌ای این است که وقتی اندازه‌ی نمونه‌ای کم است، چگونه می‌توان برآوردهای معتبر به‌دست آورد؟ زمانی که اندازه‌ی نمونه‌ای به‌دست آمده از یک ناحیه، کوچک (یا حتی صفر) باشد، برآوردهای مستقیم از یک انحراف معیار بزرگ و غیر قابل قبولی برخوردار می‌شوند. یک راه بهبود این برآوردها، وام گرفتن قدرت از منبع‌های داده‌ای موجود است. برای این کار می‌توان از طریق مدل‌بندی یا به عبارتی استفاده از مدل‌های اتصالی (شامل مدل‌های صریح و ضمنی) اثر اندازه‌ی نمونه‌ای را بهبود بخشید. برای به‌دست آوردن برآورد معتبر میانگین کوچک ناحیه‌ای مدل‌های آمیخته‌ی خطی تعمیم‌یافته و بهترین پیشگوگر ناریب خطی تجربی به‌طور گسترده مورد استفاده قرار گرفته‌اند. در این مقاله ابتدا برآورد کوچک ناحیه‌ای را معرفی می‌کنیم. سپس برای به‌دست آوردن برآوردهای معتبر برای کوچک ناحیه‌ها به معرفی مدل فی-هریوت (فی و هریوت، ۱۹۷۹)، حالت خاص مدل‌های آمیخته‌ی خطی تعمیم‌یافته می‌پردازیم. سرانجام در یک مطالعه‌ی شبیه‌سازی با استفاده از داده‌های سرشماری کشاورزی سال ۱۳۸۲، تولید پرتقال در شهرستان‌های استان فارس (کوچک ناحیه‌ها) در سال ۱۳۸۲ و بر اساس مدل فی-هریوت برآورد می‌شود.

**واژه‌های کلیدی:** برآورد کوچک ناحیه‌ای، برآوردگر مستقیم، برآوردگر نامستقیم، مدل فی-هریوت، مدل‌های آمیخته‌ی خطی تعمیم‌یافته، برآورد ماکسیمم درستی، مقید، بهترین پیشگوگر ناریب خطی تجربی.

### ۱ مقدمه

این زیرجامعه‌ها، کوچک ناحیه‌ها و برآوردهای مربوط به آن‌ها، برآوردهای کوچک ناحیه‌ای<sup>۲</sup> نامیده می‌شوند. به دلیل کوچک بودن اندازه‌ی نمونه‌ای این ناحیه‌ها که ممکن است حتی گاهی صفر نیز باشد، اغلب برآوردهای مستقیم از انحراف استاندارد بزرگ و غیر قابل قبولی برخوردار هستند. یک راه بهبود دقت برآوردها، وام گرفتن قدرت<sup>۳</sup> از منبع‌های داده‌ای مرتبط (داده‌های به‌دست آمده از کوچک ناحیه‌های مشابه یا مجاور در طول زمان یا داده‌های حاصل از آخرین آمارگیری همان کوچک ناحیه یا ترکیبی از این دو) است. بنابراین وام گرفتن قدرت یعنی افزایش اثر اندازه‌ی نمونه‌ای برای برآورد که با استفاده از مدل‌هایی (صریح یا ضمنی) برای ربط دادن داده‌های اضافی و کمکی به داده‌های نمونه‌ای انجام

به‌منظور صرفه‌جویی در هزینه‌ها و زمان و نیز دقت قابل قبول آماره‌ها، واحدهای تولیدکننده‌ی آمار تمایل به تولید آمار در سطح ملی دارند. اما در طول چند دهه‌ی گذشته مدیران و مسئولان، سیاست‌گذاران و برنامه‌ریزان، درخواست‌های زیادی برای برآوردهای مربوط به زیرجامعه‌ها داشته‌اند. به‌طور معمول آمارگیری‌های نمونه‌ای برای تولید برآوردهای پارامترهای جامعه‌ی بزرگ به‌کار گرفته می‌شوند. زیرجامعه‌ها می‌توانند:

- ۱- گروه‌های اجتماعی-جمعیتی، مانند: سن، جنسیت، گروه‌های قومی - قبیله‌ای درون یک ناحیه‌ی جغرافیایی بزرگ؛ و
- ۲- ناحیه‌های جغرافیایی مانند: استان، شهرستان، شهر و ... باشند.

<sup>۱</sup>عضو هیئت علمی گروه آمار، دانشگاه علامه طباطبائی

<sup>۲</sup>small area estimations

<sup>۳</sup>borrow strength

که در آن  $y$  بردار تصادفی متغیرهای پاسخ است، ماتریس  $X_{n \times p}$  ماتریس طرح بوده که فرض می‌شود دارای رتبه‌ی کامل ستونی است و بردار  $\beta_{p \times 1}$  بردار اثرهای ثابت است. در بعضی از موارد برخی اثرهای ثابت تصادفی فرض می‌شوند (جیانگ و لاهیری، ۲۰۰۶). این حالت زمانی اتفاق می‌افتد که مشاهده‌ها به یکدیگر وابسته باشند. بردار  $e$  شامل خطاهای مدل با بردار میانگین  $0$  و ماتریس واریانس کوواریانس  $\sigma_e^2 I_n$  است. بردار  $y$  دارای توزیع  $N_n(X\beta, \sigma_e^2 I_n)$  است ( $N_n$  نمادی است که برای نشان دادن توزیع نرمال  $n$  بعدی به کار گرفته شده است). اگر به مدل رگرسیونی معمولی برداری از اثرهای تصادفی اضافه شود، مدل آمیخته‌ی خطی حاصل می‌شود که به صورت زیر است:

$$y = X\beta + Zu + e$$

که در آن ماتریس  $Z_{n \times q}$  مانند  $X$  ماتریس طرحی است. برای بردارهای  $u$  و  $e$  فرض‌های زیر وجود دارند.

$$E(u) = 0; Cov(u) = G = G(\sigma_u); u \sim N_q(0, G),$$

$$E(e) = 0; Cov(e) = R = R(\sigma_e); e \sim N_n(0, R)$$

$$Cov(u, e) = 0$$

که  $G$  و  $R$  ماتریس‌های کوواریانس متناهی دلخواه هستند. تحت فرض‌های بالا خواهیم داشت:

$$E(y) = X\beta, Cov(y) = V = V(\sigma_e) = ZGZ' + R,$$

$$y \sim N_n(X\beta, ZGZ' + R).$$

### ۱.۱.۲ برآورد پارامترهای مدل

در مدل آمیخته‌ی خطی بالا پارامترهای  $\beta$  و  $V$  و  $u$  نامعلوم هستند که باید برآورد شوند. برای برآورد بردار پارامترهای  $\beta$  و ماتریس  $V$  روش‌های مختلفی وجود دارند که در زیر به آن‌ها اشاره می‌شود. در مدل‌های آمیخته‌ی خطی برای برآورد بردار پارامتری  $\beta$  اغلب از روش کم‌ترین توان‌های دوم تعمیم‌یافته ( $GLS$ )<sup>۴</sup> استفاده می‌شود که از معادله‌های نرمال  $X'V^{-1}X\beta = X'V^{-1}y$  به دست می‌آید و برابر است با:

$$\hat{\beta}_{GLS} = (X'V^{-1}X)^{-1}X'V^{-1}y. \quad (1)$$

می‌شود (رائو، ۲۰۰۳). برآوردهایی که از این روش به دست می‌آیند برآوردهای نامستقیم نامیده می‌شوند. برای محاسبه‌ی برآوردهای نامستقیم، در دسترس بودن داده‌های کمکی خوب و تعیین مدل‌های ربط‌دهنده‌ی مناسب بسیار مهم هستند. یک مدل مناسب، برای زمانیکه اندازه‌ی نمونه‌ای کافی نیست و یا اطلاعات لازم در دسترس نیستند، مدل آمیخته‌ی خطی تعمیم‌یافته است. از جمله روش‌های برآورد نامستقیم بر اساس مدل‌های ربط‌دهنده‌ی صریح در سطح ناحیه مدل فی-هریوت است که اخیراً بسیار مورد توجه قرار گرفته است.

## ۲ کاربرد مدل‌های آمیخته‌ی خطی در برآوردهای کوچک ناحیه‌ای

در طول سه دهه‌ی گذشته، مدل‌های آمیخته‌ی خطی، در سطح وسیعی از کاربردهای کوچک ناحیه‌ای به کار گرفته شده‌اند. چنین مدل‌هایی، انعطاف‌پذیری زیادی در ترکیب اطلاعات از منبع‌های مختلف دارند. یکی از کاربردهای مهم مدل‌های آمیخته‌ی خطی حل مشکل برآورد کوچک ناحیه‌ای است، به‌ویژه زمانی که اندازه‌ی نمونه‌ای به دست آمده در کوچک ناحیه‌ها کافی نباشند. با استفاده از مدل‌های آمیخته‌ی خطی می‌توان اطلاعات دو یا چند کوچک ناحیه را با هم ادغام کرد تا یک نمونه‌ی مناسب حاصل شود. با استفاده از بهترین پیشگوگر نارایب خطی تجربی در مدل‌های آمیخته‌ی خطی می‌توان برآوردی معتبر برای کوچک ناحیه‌ها به دست آورد. در این بخش مدل‌های آمیخته‌ی خطی تعمیم‌یافته معرفی می‌شوند. سپس برآورد پارامترهای نامعلوم این مدل‌ها شامل بردار پارامتری  $\beta$  و واریانس مدل و بهترین پیشگوگر نارایب خطی تجربی مورد بررسی قرار می‌گیرند. همچنین کاربرد مدل فی-هریوت در برآورد کوچک ناحیه‌ای تشریح می‌شود.

### ۱.۲ مدل‌های آمیخته‌ی خطی تعمیم‌یافته (GLMM)

مدل خطی که فقط شامل اثرهای ثابت است را در نظر بگیرید،

$$y = X\beta + e$$

<sup>۴</sup>generalized least square

می‌شود. سیرل و همکاران (۱۹۹۲) معادله‌ی برآورد  $ML$  برای واریانس را به دست آورده‌اند. در کل، ماکسیم‌سازی به استفاده از روش‌های عددی مانند الگوریتم نیوتن-رافسون، امتیاز فیشری و الگوریتم  $EM$  نیاز دارد.

## ۲.۲ مدل فی-هریوت

فی و هریوت اطلاعات کمکی از دیگر منبع‌های داده‌ای مانند داده‌های ثبتي را برای کاهش میانگین توان دوم خطای ( $MSE$ ) برآوردگرهای کوچک ناحیه‌ای به کار گرفتند. ویژگی‌های برآورد کوچک ناحیه‌ای مانند اریبی و  $MSE$  به طور شرطی از اطلاعات کمکی نتیجه می‌شوند و فرض می‌شود که این اطلاعات برای همه‌ی ناحیه‌ها در دسترس‌اند و بدون خطا، اندازه‌گیری می‌شوند. مدل فی-هریوت یک مدل پایه‌ای در سطح ناحیه است. در این مدل واحدهای مشاهده‌ای برای هر دو متغیر پاسخ و کمکی ناحیه‌ها هستند. این مدل به صورت زیر است:

$$y_i = X_i' \beta + u_i + \varepsilon_i = \theta_i + \varepsilon_i; i = 1, \dots, m.$$

که  $\theta_i = X_i' \beta + u_i$  میانگین کوچک ناحیه‌ی  $i$ ام و غیرقابل مشاهده و  $m$  تعداد کوچک ناحیه‌ها است. بردار  $X_i$  یک بردار  $1 \times p$  از متغیرهای تبیینی و  $\beta$  یک بردار  $1 \times p$  از ضریب‌های رگرسیونی و  $\varepsilon_i$  خطای طرح و  $u_i$  اثر تصادفی خاص کوچک ناحیه‌ی  $i$ ام است و همان خطای مدل هستند که به طور مستقل از هم توزیع شده و دارای توزیع‌های زیر هستند:

$$u_i \sim N(0, A); \varepsilon_i \sim N(0, D_i); Cov(u_i, \varepsilon_i) = 0$$

$$y_i \sim N(X_i' \beta, V_i); V_i = A + D_i.$$

که  $A$  واریانس خطای مدل و نامعلوم است و  $D_i$  واریانس خطای نمونه‌گیری و معلوم فرض می‌شود. با این وجود گاهی قابل قبول است که  $D_i = \frac{\sigma_e^2}{n_i}$  در نظر گرفته شود. در اینجا  $n_i$  اندازه نمونه‌ای برای ناحیه‌ی  $i$ ام و  $\sigma_e^2$  تغییرات درون ناحیه‌ای است. در ادامه به پیشگویی ترکیب خطی اثرهای آمیخته تحت مدل فی-هریوت می‌پردازیم که کاربردهای زیادی در برآورد کوچک ناحیه‌ای دارد.

<sup>۰</sup>variance components

<sup>۱</sup>model-dependent

<sup>۲</sup>restricted maximum likelihood estimation

در مدل‌های آمیخته‌ی خطی تعمیم‌یافته  $\hat{\beta}_{GLS}$  بهترین برآوردگر ناریب خطی برای  $\beta$  است. اگر ماتریس واریانس کوواریانس در رابطه‌ی (۱) نامعلوم باشد، برآورد آن جایگزین می‌شود.

یک موضوع مهم در تحلیل مدل‌های آمیخته‌ی خطی، برآورد مؤلفه‌های واریانس<sup>۰</sup> است. اما برای ارزیابی تغییرات برآوردگرها و پیشگوگرها لازم است که کمیت‌های دیگری نیز مانند اثرهای تصادفی و آمیخته برآورد شوند. برای مدل‌ها با مؤلفه واریانس روش کلاسیک برآورد واریانس، روش تحلیل واریانس ( $ANOVA$ ) است. همچنین دو روش برآورد که به طور فراوان برای برآورد مؤلفه‌های واریانس به کار می‌روند روش‌های ماکسیم درست‌نمایی ( $ML$ ) و ماکسیم درست‌نمایی مقید ( $REML$ ) هستند. در برآورد پارامترهای واریانس به روش ماکسیم درست‌نمایی مشکل‌هایی وجود دارد. اول این که برآورد واریانس‌های به دست آمده از این روش گرایش به اریبی دارند. مشکل دیگر برآوردگرهای  $ML$ ، مدل وابسته<sup>۱</sup> بودن آن‌ها است که نسبت به ماتریس مدل بسیار حساس‌اند (دیگل و همکاران، ۲۰۰۲). به خاطر وجود چنین مشکل‌هایی، روش برآورد درست‌نمایی دیگری برای برآورد پارامترهای واریانس وجود دارد که روش ماکسیم درست‌نمایی مقید<sup>۲</sup> نامیده می‌شود. روش برآورد ماکسیم درست‌نمایی مقید بر اساس همان اصل درست‌نمایی است و برآوردهای حاصل همان ویژگی‌ها نظیر سازگاری، کارایی و فرض نرمال بودن مجانبی برآوردگرهای  $ML$  را دارا می‌باشند. از این رو روش  $REML$  با توجه به ناریب بودن یا تقریباً ناریب بودن، در مدل‌های آمیخته‌ی خطی برای برآورد پارامترهای واریانس به روش برآورد  $ML$  ارجحیت دارد. با توجه به  $\hat{\beta}_{GLS}$ ، لگاریتم تابع درست‌نمایی برای واریانس برابر است با:

$$\begin{aligned} \text{Log}L(V) = & -\frac{n-p}{2} \text{Log}(\pi) - \frac{1}{2} \text{Log}|V| + \frac{1}{2} \text{Log}|X'X| \\ & + \frac{1}{2} \text{Log}|X'V^{-1}X| - \frac{1}{2} (r(\sigma))' V^{-1} r(\sigma) \end{aligned}$$

که در آن

$$r(\sigma) = (y - X\hat{\beta}_{GLS}) = y - X(X'V(\sigma)^{-1}X)^{-1}X'V(\sigma)^{-1}y.$$

که با ماکسیم کردن تابع لگاریتم درست‌نمایی،  $\hat{\sigma}_{REML}$  حاصل

۱.۲.۲ پیشگویی ترکیب خطی اثرهای آمیخته

برابر است با (داتا و لاهیری، ۲۰۰۰):

$$MSE(\hat{\theta}_{i,FH}) = MSE(\tilde{\theta}_{i,FH}) + E(\hat{\theta}_{i,FH} - \tilde{\theta}_{i,FH})^2 \\ = g_{1i}(A) + g_{2i}(A) + g_{3i}(A) \quad (3)$$

فرض کنید که اطلاعات کمکی در دسترس هستند. تحت مدل فی-هریوت ترکیب خطی  $\theta_i = X'_i\beta + u_i$  را در نظر می‌گیریم. اگر پارامترهای  $D_i$  و  $A$  و  $\beta$  معلوم باشند، تحت فرض‌های نرمال، بهترین پیشگوگر ناریب خطی برای  $\theta_i$  برابر است با (لوهر و یبارا، ۲۰۰۸):

$$g_{1i}(A) = \frac{AD_i}{A + D_i} = \gamma_i D_i;$$

$$g_{2i}(A) = D_i^{\gamma_i} (A + D_i)^{-\gamma_i} x'_i \left[ \sum_{u=1}^m (A + D_u)^{-1} x_u x'_u \right]^{-1} x_i;$$

$$g_{3i}(A) = \gamma_i D_i^{\gamma_i} \left[ \sum_{u=1}^m (A + D_u)^{-\gamma_i} \right]^{-1}.$$

$$\tilde{\theta}_{i,FH} = X'_i\beta + \tilde{u}_i \\ = X'_i\beta + \frac{A}{A + D_i} (y_i - X'_i\beta) \\ = \gamma_i y_i + (1 - \gamma_i) X'_i\beta,$$

روش‌های برآورد کوچک ناحیه‌ای، ترکیب برآوردهای مستقیم یک آمارگیری با پیشگویی‌هایی از یک مدل، در به‌دست آوردن کمیت‌های مورد نظر با کاهش میانگین توان دوم خطا هستند.

به‌طوری‌که  $\gamma_i = \frac{A}{A + D_i}$ . این پیشگوگر یک ترکیب محدب از برآوردگر مستقیم و مقدار  $X'_i\beta$  مدل رگرسیونی است. میانگین توان دوم خطای این پیشگوگر عبارت است از:

$$MSE(\tilde{\theta}_{i,FH}) = E(\tilde{\theta}_{i,FH} - \theta_i)^2 \\ = (\gamma_i - 1)^2 E(u_i)^2 + \gamma_i^2 E(\varepsilon_i)^2 \\ + 2(\gamma_i - 1)\gamma_i E(u_i\varepsilon_i)$$

۳ کاربرد مدل فی-هریوت در برآورد کوچک ناحیه‌ای

در این بخش با به‌کارگیری مدل فی-هریوت که یک مدل پایه‌ای در سطح ناحیه است، تولید کل محصول پرتقال در شهرستان‌های استان فارس (کوچک ناحیه‌ها) با استفاده از داده‌های سرشماری کشاورزی سال ۱۳۸۲ مرکز آمار ایران برآورد می‌شود.

با توجه به فرض استقلال  $u_i$  و  $\varepsilon_i$  و این‌که  $E(u_i) = 0$  و  $E(\varepsilon_i) = 0$  خواهیم داشت:

$$MSE(\tilde{\theta}_{i,FH}) = \frac{D_i^{\gamma_i}}{(A + D_i)^{\gamma_i}} Var(u_i) + \frac{A^{\gamma_i}}{(A + D_i)^{\gamma_i}} Var(\varepsilon_i) \\ = \frac{D_i^{\gamma_i} A}{(A + D_i)^{\gamma_i}} + \frac{A^{\gamma_i} D_i}{(A + D_i)^{\gamma_i}} \\ = \frac{AD_i}{A + D_i} = \gamma_i D_i.$$

سرشماری کشاورزی توسط مرکز آمار ایران انجام می‌شود. روش گردآوری داده‌ها از طریق پرسش‌نامه‌ای که از هر بهره‌بردار پرسیده می‌شود، صورت می‌گیرد. در طراحی پرسش‌نامه مواردی از قبیل: تعداد بهره‌برداری؛ سطح کاشت محصول (برحسب هکتار)؛ تعداد درخت بارور و تعداد درخت نهال و مقدار تولید (بر حسب تن) در نظر گرفته می‌شود. نتیجه‌ی این سرشماری حکایت از وجود ۹۲۹۸ بهره‌برداری محصول پرتقال داشته و تولید کل پرتقال ۹۷۹۶۶ تن در سطح استان بوده است.

در اصل پارامترهای  $A$  و  $\beta$  نامعلومند و باید توسط داده‌ها برآورد شوند. بنا بر این بهترین پیشگوگر ناریب خطی تجربی برای  $\theta_i$  برابر است با:

$$\hat{\theta}_{i,FH} = X'_i\hat{\beta} + \hat{u}_i, \\ = \hat{\gamma}_i y_i + (1 - \hat{\gamma}_i) X'_i\hat{\beta}, \quad (2)$$

در این کاربرد برای به‌دست آوردن برآورد پارامتر موردنظر (تولید کل محصول پرتقال در استان فارس در سال ۱۳۸۲) با در نظر گرفتن تولید کل محصول نارنگی به‌عنوان متغیر کمکی در شهرستان‌های استان فارس (کوچک ناحیه) از رویکرد مدل مینا شامل مدل صریح (مدل فی-هریوت) استفاده می‌شود.

که در آن  $\hat{\gamma}_i = \frac{\hat{A}}{\hat{A} + D_i}$ . رابطه‌ی (۲)، یک ترکیب محدب از برآوردگر مستقیم  $y_i$  و مقدار پیشگویی  $X'_i\hat{\beta}$  معادله‌ی رگرسیونی است. میانگین توان دوم خطای این پیشگویی با توجه به این‌که

$$l' = X'_i; V = A + D_i; G = A; Z = 1; m = 1,$$

## ۱.۳ تعیین اندازه‌ی نمونه‌ای

جدول ۱: اندازه‌ی نمونه‌ای قرار گرفته در شهرستان‌های استان

فارس از نمونه‌ی کل

| ردیف | شهرستان    | اندازه‌ی نمونه‌ای ( $n_i$ ) | واریانس ( $s_i^2$ ) |
|------|------------|-----------------------------|---------------------|
| ۱    | استهبان    | ۴۰                          | ۳۵/۸۲               |
| ۲    | جهرم       | ۶۸۹                         | ۱۶۳/۲۵              |
| ۳    | داراب      | ۵۵۵                         | ۴۵۴/۱۸              |
| ۴    | زرین دشت   | ۳                           | ۰/۱۳                |
| ۵    | شیراز      | ۶                           | ۰/۲۶                |
| ۶    | فراشبند    | ۷                           | ۰/۰۰۹               |
| ۷    | فسا        | ۱۵۰                         | ۳۵۲۱/۷۸             |
| ۸    | فیروزآباد  | ۴۰                          | ۰/۴۵                |
| ۹    | قیروکارزین | ۸۰                          | ۶۱۶/۳۹              |
| ۱۰   | کازرون     | ۱۶۰                         | ۳۰۷/۲۷              |
| ۱۱   | لارستان    | ۳۵                          | ۴۸۳/۵۷              |
| ۱۲   | لامرد      | ۳                           | ۰/۰۵۶               |
| ۱۳   | مرودشت     | ۵                           | ۰/۰۵۱               |
| ۱۴   | ممسنی      | ۱۱۵                         | ۵۰/۴۶               |
| ۱۵   | مهر        | ۳                           | ۰/۰۸                |
| ۱۶   | نیریز      | ۹                           | ۲/۶۱                |

برای به‌دست آوردن برآورد کوچک ناحیه‌ای از داده‌های آمارگیری نمونه‌ای که برای اجرا در ناحیه‌ی بزرگ‌تر شامل کوچک ناحیه‌ها طراحی شده است، استفاده می‌شود. نمونه‌های قرار گرفته در هر شهرستان به عنوان اندازه‌ی نمونه‌ای هر شهرستان (کوچک ناحیه‌ها) در نظر گرفته می‌شود.

برای به‌دست آوردن اندازه‌ی نمونه‌ای کل،  $n$ ، در سطح استان، با استفاده از یک نمونه‌ی تصادفی به تعداد ۲۰۰ بهره‌برداری از فهرست حاصل از سرشماری کشاورزی سال ۱۳۸۲ واریانس مجموع تولید در سطح استان را برآورد و سپس از فرمول زیر مقدار  $n$  را برای  $\alpha = ۰/۰۵$  محاسبه کردیم.

$$n = \frac{Z_{\alpha/2}^2 \sum s_i^2 N_i^2}{e^2 + Z_{\alpha/2}^2 \sum s_i^2 N_i} \quad (۴)$$

در رابطه‌ی (۴)،  $Z_{\alpha/2} \approx ۲$ ،  $N=۹۲۹۸$  و  $s^2$  برآورد واریانس مجموع است که برابر با  $۹۴۸/۱۱۶$  به‌دست آمده و  $e$  حاشیه‌ی خطا است که در اینجا بر اساس  $۰/۱۲$  از تولید کل محصول پرتقال ( $۰/۱۲ \times ۹۷۹۶۶ = ۱۱۷۵۵/۹۲$ ) در سطح استان در نظر گرفته شده است. اندازه‌ی نمونه‌ای برای یک نمونه‌گیری در سطح استان فارس در زیر محاسبه شده است.

$$n = \frac{۴ \times ۹۴۸/۱۱۶ \times (۹۲۹۸)^2}{(۱۱۷۵۵/۹۲)^2 + ۴ \times ۹۴۸/۱۱۶ \times (۹۲۹۸)} \approx ۱۹۰۰$$

پس از انتخاب این تعداد نمونه در سطح استان فارس، تعداد نمونه‌های قرار گرفته شده در شهرستان‌های این استان محاسبه و در جدول ۱ ارائه شده است.

برای تعیین اندازه‌ی نمونه‌ای برای هر شهرستان، با استفاده از یک نمونه‌ی تصادفی مقدماتی از هر شهرستان و با به‌کارگیری رابطه‌ی (۴) و با در نظر گرفتن  $\alpha = ۰/۰۵$  و هم‌چنین با در نظر گرفتن حاشیه‌ی خطای  $۰/۱۲$  از تولید کل محصول پرتقال برای شهرستان داراب و  $۰/۱۵$  از تولید کل محصول پرتقال در بقیه‌ی شهرستان‌ها، اندازه‌ی نمونه‌ای برای شهرستان‌های استان فارس به‌دست آمد. این اندازه‌های نمونه‌ای در جدول ۲ آمده‌اند.

جدول ۲: اندازه‌ی نمونه‌ای بهین محاسبه شده  $n_i$  برای

شهرستان‌های استان فارس

| ردیف | شهرستان    | تعداد بهره‌برداری $N_i$ | حاشیه‌ی خطا $e_i$ | اندازه‌ی نمونه‌ای مقدماتی | $n_i$ |
|------|------------|-------------------------|-------------------|---------------------------|-------|
| ۱    | استهبان    | ۱۵۳                     | ۱۶۲/۴۵            | ۱۵                        | ۷۸    |
| ۲    | جهرم       | ۳۴۶۷                    | ۲۳۱۴/۸            | ۵۰                        | ۱۲۹۹  |
| ۳    | داراب      | ۳۰۰۳                    | ۷۸۵۲/۵            | ۵۰                        | ۱۶۶۰  |
| ۴    | زرین دشت   | ۸                       | ۰/۳               | ۵                         | ۶     |
| ۵    | شیراز      | ۲۸                      | ۲/۱               | ۸                         | ۴۷    |
| ۶    | فراشبند    | ۳۱                      | ۲/۸۵              | ۵                         | ۵۱    |
| ۷    | فسا        | ۷۴۵                     | ۲۳۷۶/۴۵           | ۴۰                        | ۶۹۴   |
| ۸    | فیروزآباد  | ۱۵۲                     | ۱۴/۷              | ۲۰                        | ۱۶۰   |
| ۹    | قیروکارزین | ۲۸۸                     | ۳۸۸/۲             | ۳۰                        | ۳۱۶   |
| ۱۰   | کازرون     | ۷۰۰                     | ۱۳۲۰/۱۵           | ۳۵                        | ۷۱۸   |
| ۱۱   | لارستان    | ۱۴۷                     | ۱۵۹/۱۵            | ۲۰                        | ۴۰    |
| ۱۲   | لامرد      | ۹                       | ۰/۳               | ۵                         | ۱۹    |
| ۱۳   | مرودشت     | ۱۴                      | ۰/۳               | ۵                         | ۲۳    |
| ۱۴   | ممسنی      | ۴۹۹                     | ۹۲/۲۵             | ۳۰                        | ۴۹۳   |
| ۱۵   | مهر        | ۱۶                      | ۱/۰۵              | ۵                         | ۱۷    |
| ۱۶   | نیریز      | ۴۰                      | ۱۸/۳              | ۸                         | ۳۴    |

با مقایسه‌ی جدول‌های ۱ و ۲ با یکدیگر، نتیجه می‌شود که اندازه‌ی نمونه‌ای قرار گرفته در هر شهرستان وقتی اندازه‌ی نمونه‌ای در سطح استان محاسبه شده است، کوچک‌تر از اندازه‌ی نمونه‌ای بهین محاسبه شده برای هر شهرستان است. واریانس هر شهرستان در جدول ۱ به دلیل کوچک بودن اندازه‌ی نمونه‌ای، بزرگ است. بنا بر این در این حالت این شهرستان‌ها کوچک ناحیه به‌شمار می‌آیند و برآوردهای مستقیم برای برآورد مجموع در این کوچک ناحیه‌ها نادقیق‌اند لذا باید از فن‌های برآورد کوچک ناحیه‌ای برای بهبود برآوردها استفاده شود.

### ۲.۳ برآورد تولید کل محصول پرتقال

برای به‌دست آوردن برآورد تولید کل محصول پرتقال در هر شهرستان (کوچک ناحیه‌ها) با به‌کارگیری رویکرد مدل‌مبنا، باید یک مدل مناسب به داده‌ها برازش داده شود. با استفاده از منبع داده‌ای موجود (سرشماری کشاورزی ۱۳۸۲)، تولید نارنگی در هر شهرستان را به عنوان متغیر کمکی در نظر گرفتیم. از آن‌جایی که مجموع تولید نارنگی در سطح هر شهرستان از سرشماری کشاورزی سال ۱۳۸۲ در دسترس است، لذا می‌توانیم از مدل فی-هریوت که در آن واحدهای مشاهده‌ای برای هر دو متغیر پاسخ و کمکی در سطح ناحیه‌ها هستند، برای برآورد تولید کل محصول پرتقال استفاده کنیم. همان‌طور که در بخش ۳ اشاره شد، مدل فی-هریوت به صورت زیر است:

$$y_i = X_i' \beta + u_i + \varepsilon_i = \theta_i + \varepsilon_i; i = 1, \dots, 16.$$

نمادها در بخش ۲ معرفی شده‌اند. با توجه به این‌که به ازای هر  $y_i$  فقط یک متغیر کمکی داریم، در نتیجه مدل فی-هریوت، یک مدل فی-هریوت تک متغیری است. مقدار  $y_i$  ناحیه‌ای و  $X_i$  برای هر شهرستان در جدول ۳ آورده شده است:

با توجه به این‌که هدف، برآورد تولید کل محصول پرتقال در شهرستان‌های استان فارس است، تحت مدل فی-هریوت ترکیب خطی اثر آمیخته‌ی  $\theta_i = X_i' \beta + u_i$  را در نظر می‌گیریم. بهترین پیشگوگر ناریب خطی تجربی برای  $\theta_i$  برابر است با:

$$\hat{\theta}_{i, FH} = \hat{\gamma}_i y_i + (1 - \hat{\gamma}_i) X_i' \hat{\beta}(\theta)$$

که در آن  $\hat{\gamma}_i = \frac{\hat{A}}{\hat{A} + \hat{D}_i}$ ،  $\hat{\beta} = (X' \hat{V}^{-1} X)^{-1} X' \hat{V}^{-1} y$ ،  $\hat{D}_i = \frac{N-n}{nN} s_i^2$  و  $X' = (X_1, \dots, X_{16})$ ،  $\hat{V} = \hat{A} + \hat{D}_i$

جدول ۳: مقدارهای محصول پرتقال،  $y_i$ ، و نارنگی،  $X_i$ ، بر

حسب تن

| ردیف | شهرستان    | $y_i$   | $X_i$ |
|------|------------|---------|-------|
| ۱    | استهبان    | ۲۸۷/۵۳  | ۶     |
| ۲    | جهرم       | ۲۸۲۷/۸  | ۱۳۷۴۰ |
| ۳    | داراب      | ۸۸۵۳/۰۴ | ۷۹۹   |
| ۴    | زرین دشت   | ۰/۷۴    | ۰/۵   |
| ۵    | شیراز      | ۱/۸۲    | ۶     |
| ۶    | فراشبند    | ۲/۲     | ۰/۵   |
| ۷    | فسا        | ۴۷۹۲/۶  | ۷۱۰۶  |
| ۸    | فیروزآباد  | ۱۲/۱۱۵  | ۹۷    |
| ۹    | قیروکارزین | ۸۴۸     | ۱۹۹۵  |
| ۱۰   | کازرون     | ۱۶۰۷/۲۷ | ۵۵۳۵  |
| ۱۱   | لارستان    | ۳۵۶/۷۵  | ۳۸۷   |
| ۱۲   | لامرد      | ۰/۵۳    | ۱     |
| ۱۳   | مرودشت     | ۰/۳۷    | ۱     |
| ۱۴   | ممسنی      | ۱۵۵/۵۷  | ۱۰۴   |
| ۱۵   | مهر        | ۰/۵۵    | ۲     |
| ۱۶   | نیریز      | ۲۵/۹۴   | ۴     |

با استفاده از فرمول (۵)،  $\hat{\beta} = ۲/۱$  به‌دست آمد که در سطح ۵ درصد معنی‌دار است. همچنین بازه‌ی اطمینان تقریبی برای برابر با (۲/۸۱ و ۱/۳۱) به‌دست آمد که شامل صفر نیست. برای بررسی فرض نرمال بودن  $e_i$ ها (برآورد  $\varepsilon_i$ ها) در سطح  $\alpha = ۰/۰۵$  آزمون کولموگوروف-اسمیرونوف انجام شد. تحت این آزمون  $p$  مقدار آزمون برابر با ۰/۲۱ شد لذا فرض نرمال بودن  $e_i$ ها رد نمی‌شود. برای به‌دست آوردن برآورد واریانس اثرهای تصادفی،  $\hat{A}$ ، از روش برآوردگر گشتاوری ساده استفاده شد. برآورد این واریانس برابر است با:

$$\hat{A} = \frac{1}{m-p} \left\{ \sum_i (Z_i^{-1} y_i - X_i' \hat{\beta})^2 - \sum_i \frac{D_i}{Z_i^2} (1 - \tilde{h}_{ii}) \right\} = ۱۷/۷۱$$

که در آن  $m = ۱۶$  (برابر با تعداد کوچک ناحیه‌ها)،  $p = ۱$  (تعداد پارامترهای برآورد شده،  $\hat{\beta}$ ) و  $Z_i = ۱$  و  $\hat{\theta}_{i, FH} = X_i' (\sum_I X_i X_i')^{-1} X_i$  بنا بر این  $\hat{\gamma}_i$  و  $\hat{V}_i$  و  $\hat{D}_i$

برای هر شهرستان با توجه به رابطه‌ی (۵) و همچنین مقدار  $y_i$  و  $X_i$  برای هر شهرستان محاسبه و در جدول ۴ آورده شده‌اند:

با به‌دست آوردن میانگین توان دوم خطای برآوردهای کوچک ناحیه‌ای به‌دست آمده می‌توان دقت و اعتبار این برآوردها را ارزیابی کرد. با توجه به رابطه‌ی (۳) میانگین توان دوم خطا برابر است با:

جدول ۴: مقدار پارامترهای برآورد شده‌ی  $\hat{\theta}_{i,FH}$ ،  $\hat{\gamma}_i$ ،  $\hat{V}_i$ ،  $\hat{D}_i$

$$\widehat{MSE}(\hat{\theta}_{i,FH}) = g_{1i}(\hat{A}) + g_{2i}(\hat{A}) + g_{3i}(\hat{A})$$

که در آن

$$g_{1i}(\hat{A}) = \frac{\hat{A}\hat{D}_i}{\hat{A} + \hat{D}_i} = \hat{\gamma}_i\hat{D}_i;$$

$$g_{2i}(\hat{A}) = \hat{D}_i^2(\hat{A} + \hat{D}_i)^{-2}x'_i[\sum_{u=1}^m(\hat{A} + \hat{D}_u)^{-1}x_u x'_u]^{-1}x_i;$$

$$g_{3i}(\hat{A}) = 2\hat{D}_i^2(\hat{A} + \hat{D}_i)^{-2}[\sum_{u=1}^m(\hat{A} + \hat{D}_u)^{-2}]^{-1}.$$

و همچنین  $\widehat{MSE}(\hat{Y}_{i,FH}) = N_i \widehat{MSE}(\hat{\theta}_{i,FH})$  که  $N_i$  تعداد بهره‌برداری در شهرستان نام است. مقدارهای تقریبی میانگین توان دوم خطا برای برآوردهای به‌دست آمده در جدول ۶ آورده شده‌اند (مقدار  $\widehat{MSE}(\hat{Y}_{i,FH})$  در مقیاس ۰/۰۰۰۱ تولید بیان شده است). بر این اساس برآوردهای هر بیوت تولید کل محصول پرتقال برای هر شهرستان برابر با  $\hat{Y}_{i,FH} = N_i \hat{\theta}_{i,FH}$  است که مقدار آن برای هر شهرستان (کوچک ناحیه) در جدول ۵ آورده شده است.

| ردیف | شهرستان    | $\hat{D}_i$ | $\hat{V}_i$ | $\hat{\gamma}_i$ | $\hat{\theta}_{i,FH}$ |
|------|------------|-------------|-------------|------------------|-----------------------|
| ۱    | استهبان    | ۰/۶۶        | ۱۸/۳۷       | ۰/۹۶             | ۶/۹۲                  |
| ۲    | جهرم       | ۰/۱۶        | ۱۷/۹        | ۰/۹۹             | ۴/۱۸                  |
| ۳    | داراب      | ۰/۶۷        | ۱۸/۳۷       | ۰/۹۶             | ۱۵/۷                  |
| ۴    | زرین دشت   | ۰/۰۳        | ۱۷/۷۴       | ۰/۹۹             | ۰/۲۸                  |
| ۵    | شیراز      | ۰/۳         | ۱۷/۷۴       | ۰/۹۹             | ۰/۳                   |
| ۶    | فراشبند    | ۰/۰۰۱       | ۱۷/۷۱       | ۰/۹۹۹            | ۰/۳                   |
| ۷    | فسا        | ۱۸/۷۵       | ۳۶/۴۶       | ۰/۴۸             | ۲۱/۵۲                 |
| ۸    | فیروزآباد  | ۰/۰۱        | ۱۷/۷۲       | ۰/۹۹۹            | ۰/۳۰۱                 |
| ۹    | قیروکارزین | ۵/۵۶        | ۲۳/۲۷       | ۰/۷۶             | ۹/۹۶                  |
| ۱۰   | کازرون     | ۴۸/۱        | ۱۹/۱۹       | ۰/۹۲             | ۱۰/۸۷                 |
| ۱۱   | لارستان    | ۱۰/۵۳       | ۲۸/۲۴       | ۰/۶۳             | ۹/۰۲                  |
| ۱۲   | لامرد      | ۰/۱۲        | ۱۷/۷۲       | ۰/۹۹۹            | ۰/۱۷۵                 |
| ۱۳   | مرودشت     | ۰/۰۰۱       | ۱۷/۷۱       | ۰/۹۹۹            | ۰/۰۷۴                 |
| ۱۴   | ممسنی      | ۰/۳۴        | ۱۸/۰۵       | ۰/۹۸             | ۱/۳۴                  |
| ۱۵   | مهر        | ۰/۰۲۲       | ۱۷/۷۳       | ۰/۹۹۸            | ۰/۱۸                  |
| ۱۶   | نیریز      | ۰/۲۲        | ۱۷/۹۳       | ۰/۹۹             | ۲/۸۶                  |

جدول ۵: برآورد فی-هریوت  $\hat{Y}_{i,FH}$  برآورد مستقیم،  $\hat{Y}_i$  و مقدار واقعی،  $Y_i$ ، حاصل از سرشماری کشاورزی سال ۱۳۸۲

| ردیف | شهرستان    | $\hat{Y}_{i,FH}$ | $\hat{Y}_i$ | $Y_i$ |
|------|------------|------------------|-------------|-------|
| ۱    | استهبان    | ۱۰۶۰             | ۱۰۹۹/۱۱     | ۱۰۸۳  |
| ۲    | جهرم       | ۱۴۴۸۳            | ۱۴۲۲۹/۳     | ۱۵۳۴۲ |
| ۳    | داراب      | ۴۷۱۴۳            | ۴۷۹۰۲/۱۲    | ۵۲۳۵۰ |
| ۴    | زرین دشت   | ۲/۲۲             | ۲/۲۴        | ۲     |
| ۵    | شیراز      | ۸/۳۲             | ۸/۴۹        | ۱۴    |
| ۶    | فراشبند    | ۹/۷              | ۹/۷         | ۱۹    |
| ۷    | فسا        | ۱۶۰۳۲            | ۱۹۳۷۰       | ۱۵۸۴۳ |
| ۸    | فیروزآباد  | ۴۶               | ۴۶/۰۴       | ۹۸    |
| ۹    | قیروکارزین | ۲۸۶۷/۳۳          | ۳۰۵۲        | ۲۵۸۸  |
| ۱۰   | کازرون     | ۷۶۰۹             | ۷۰۳۱/۸۲     | ۸۸۰۱  |
| ۱۱   | لارستان    | ۱۳۲۵             | ۱۴۹۸/۴      | ۱۰۶۱  |
| ۱۲   | لامرد      | ۱/۶              | ۱/۶         | ۲     |
| ۱۳   | مرودشت     | ۱/۰۴             | ۱/۰۴        | ۲     |
| ۱۴   | ممسنی      | ۶۴۴              | ۶۷۲/۳۳      | ۶۳۵   |
| ۱۵   | مهر        | ۲/۸۸             | ۲/۹۳        | ۷     |
| ۱۶   | نیریز      | ۱۱۴/۴            | ۱۱۵/۳       | ۱۲۲   |

## ۴ مطالعه‌ی شبیه‌سازی

در این مطالعه با استفاده از نمونه‌ی ۱۹۰۰ تایی انتخاب شده، ۱۰۰۰ نمونه‌ی خودگردان تولید شده و مقدار برآوردهای تولید کل و  $MSSE$  آن‌ها برای برآورد مستقیم و فی-هریوت محاسبه و مقایسه می‌شوند.

### ۱.۴ فرمول‌های برآورد

فرمول‌های به‌کار رفته برای برآورد خودگردان تولید کل محصول پرتقال برای برآوردهای مستقیم،  $\hat{Y}_i$  و فی-هریوت،  $\hat{Y}_{i,FH}$ ، بر اساس ۱۰۰۰ نمونه‌ی خودگردان به صورت زیر است:

$$\hat{Y}_{i,FH} = \frac{1}{1000} \sum_{u=1}^{1000} [N_i (\hat{\gamma}_{iu} y_{iu} + (1 - \hat{\gamma}_{iu}) X'_{iu} \hat{\beta})];$$

$$i = 1, \dots, 16 \quad (6)$$

جدول ۶: میانگین توان دوم خطای برآورد محصول پرتقال در شهرستان‌های استان فارس در سال ۱۳۸۲

| ردیف | شهرستان    | $g_{1i}(A)$ | $g_{2i}(A)$ | $g_{3i}(A)$ | $\widehat{MSE}(\hat{\theta}_{i,FH})$ | $\widehat{MSE}(\hat{Y}_{i,FH})$ |
|------|------------|-------------|-------------|-------------|--------------------------------------|---------------------------------|
| ۱    | استهبان    | ۰/۶۳        | ۰/۰۰۰۱      | ۰/۰۰۰۴      | ۰/۶۴                                 | ۱/۴۸                            |
| ۲    | جهرم       | ۰/۱۹        | ۰/۰۰۰۲      | ۰/۰۰۰۳      | ۰/۱۹                                 | ۲۲۸/۳۸                          |
| ۳    | داراب      | ۰/۶۴        | ۰/۰۰۰۳      | ۰/۰۰۰۴      | ۰/۶۵                                 | ۵۸۶/۱۷                          |
| ۴    | زرین دشت   | ۰/۰۳        | ۰           | ۰/۰۰۰۱      | ۰/۰۳                                 | ۰/۰۰۰۲                          |
| ۵    | شیراز      | ۰/۰۳        | ۰           | ۰/۰۰۰۱      | ۰/۰۳۹                                | ۰/۰۰۳                           |
| ۶    | فراشبند    | ۰/۰۰۱       | ۰           | ۰           | ۰/۰۰۱                                | ۰/۰۰۰۱                          |
| ۷    | فسا        | ۹           | ۲/۸۱        | ۰/۴۵        | ۱۲/۲۵                                | ۶۷۹/۹                           |
| ۸    | فیروزآباد  | ۰/۰۱        | ۰           | ۰           | ۰/۰۱                                 | ۰/۰۲                            |
| ۹    | قیروکارزین | ۴/۲۲        | ۰/۲۷        | ۰/۱۲        | ۴/۶۱                                 | ۲۸/۲۴                           |
| ۱۰   | کازرون     | ۱/۳۶        | ۰/۰۳۵       | ۰/۰۱        | ۱/۴                                  | ۶۸/۶                            |
| ۱۱   | لارستان    | ۶/۶۳        | ۰/۱۱۸       | ۰/۱۸        | ۶/۹۳                                 | ۱۴/۹۷                           |
| ۱۲   | لامرد      | ۰/۰۱۲       | ۰           | ۰           | ۰/۰۱۲                                | ۰/۰۰۰۱                          |
| ۱۳   | مرودشت     | ۰/۰۰۱       | ۰           | ۰           | ۰/۰۰۱                                | ۰/۰۰۰۰۲                         |
| ۱۴   | ممسنی      | ۰/۳۳        | ۰           | ۰/۰۰۱       | ۰/۳۳                                 | ۸/۱۵                            |
| ۱۵   | مهر        | ۰/۰۲۲       | ۰           | ۰           | ۰/۰۲۲                                | ۰/۰۰۰۰۶                         |
| ۱۶   | نیریز      | ۰/۲۱۸       | ۰           | ۰/۰۰۱       | ۰/۲۲                                 | ۰/۰۳۵                           |

### ۲.۴ یافته‌های مطالعه

برآورد تولید کل محصول پرتقال بر اساس روش‌های برآورد مستقیم و فی-هریوت با ۱۰۰۰ نمونه‌ی خودگردان طبق رابطه‌های (۶) و (۷) محاسبه و در جدول ۷ ارائه شده است.

به‌منظور مقایسه‌ی عمل‌کرد برآوردگرهای مستقیم و فی-هریوت، برآوردهای  $\widehat{MSE}(\hat{Y}_{i,FH})$  و  $\widehat{Var}(\hat{Y}_i)$  (در مقیاس ۰/۰۰۰۱ تولید) بر اساس ۱۰۰۰ نمونه‌ی خودگردان، در جدول ۸ آورده شده‌اند.

با توجه به جدول ۸ مشاهده می‌شود که در اغلب شهرستان‌ها، برآوردگر فی-هریوت برای برآورد تولید کل پرتقال برای ۱۶ شهرستان استان فارس عمل‌کرد بهتری نسبت به برآوردگر مستقیم داشته است. همچنین با مقایسه‌ی برآورد  $MSE$ ی دو برآوردگر مستقیم و فی-هریوت در جدول ۸ مشاهده می‌شود که برآوردگرهای مستقیم دارای برآورد  $MSE$ ی بزرگ‌تر و در نتیجه دارای عدم دقت بیشتری نسبت به برآوردهای فی-هریوت هستند.

$$\bar{Y}_i = \frac{1}{1000} \sum_{u=1}^{1000} \left[ \sum_{j=1}^{n_{iu}} \frac{N_i}{n_{iu}} y_{iju} \right];$$

$$i = 1, \dots, 16 \tag{7}$$

فرمول‌های به‌کار رفته برای برآوردهای خودگردان میانگین توان دوم خطای برآوردگرهای فی-هریوت  $\widehat{MSE}(\hat{Y}_{i,FH})$  و مستقیم  $\widehat{Var}(\hat{Y}_i)$  طبق رابطه‌ی (۳) بر اساس ۱۰۰۰ نمونه‌ی خودگردان به صورت زیر است:

$$\widehat{MSE}(\bar{Y}_{i,FH}) = \frac{1}{1000} \sum_{u=1}^{1000} \left[ N_i \widehat{MSE}(\hat{\theta}_{i,FH}) \right];$$

$$i = 1, \dots, 16$$

$$\widehat{Var}(\bar{Y}_i) = \frac{1}{1000} \sum_{u=1}^{1000} \left[ \sum_{j=1}^{n_{iu}} \frac{N_i}{n_{iu}} \widehat{Var}(\hat{Y}_{iju}) \right];$$

$$i = 1, \dots, 16$$



## ۵ نتیجه گیری

در این مقاله برای برآورد تولید کل محصول پرتقال در شهرستان‌های استان فارس (کوچک ناحیه‌ها) از رویکرد مدل مبنا (مدل فی-هریوت) استفاده شد. موضوع اساسی در برآورد کوچک ناحیه‌ای، به‌دست آوردن برآورد معتبر برای کوچک ناحیه‌ها و ارزیابی اعتبار این برآوردها از طریق میانگین توان دوم خطای برآوردگر است. در مطالعه‌ی شبیه‌سازی نشان داده شد که برآوردهای فی-هریوت به‌دلیل دارا بودن  $MSE$ ی کوچک‌تر و در نتیجه دقت بیشتر نسبت به برآوردهای مستقیم، قابل قبول‌تر از برآوردهای مستقیم برای برآورد تولید پرتقال در سطح شهرستان‌های استان فارس در سال ۱۳۸۲ است.

لازم به ذکر است که برای به‌کارگیری مدل مناسب توجه به رابطه متغیر پاسخ و متغیر کمکی مهم است. چرا که اگر مدل انتخابی مناسب نباشد ارزیابی بزرگی به وجود می‌آورد. هر چند برآوردهای مدل مبنا گرایش به ارزیابی دارند ولی در قیاس با برآوردهای طرح مبنا (برآوردگر مستقیم) واریانس‌های کوچک‌تری در زمینه‌ی برآوردهای کوچک ناحیه‌ای دارند. همچنین به‌کارگیری مدل صریح به‌دلیل اثر دادن تغییرات بین ناحیه‌ای در برآورد پارامتر موردنظر، بهبود قابل توجه‌ای در برآورد پارامترها ایجاد می‌کند.

جدول ۷: برآورد فی-هریوت  $\hat{Y}_{i,FH}$  برآورد مستقیم،  $\hat{Y}_i$  و مقدار واقعی،  $Y_i$ ، حاصل از سرشماری کشاورزی سال ۱۳۸۲

| ردیف | شهرستان    | $\hat{Y}_{i,FH}$ | $\hat{Y}_i$ | $Y_i$ |
|------|------------|------------------|-------------|-------|
| ۱    | استهبان    | ۱۱۰۵             | ۱۱۹۵        | ۱۰۸۳  |
| ۲    | جهرم       | ۱۵۲۴۰            | ۱۵۰۵۷       | ۱۵۳۴۲ |
| ۳    | داراب      | ۴۷۵۷۹            | ۴۹۴۸۰       | ۵۲۳۵۰ |
| ۴    | زرین دشت   | ۱/۹              | ۲/۱۱        | ۲     |
| ۵    | شیراز      | ۱۰/۶             | ۱۰/۲        | ۱۴    |
| ۶    | فراشبند    | ۱۱/۲۲            | ۱۱/۳        | ۱۹    |
| ۷    | فسا        | ۱۶۰۹۶            | ۱۷۲۲۵       | ۱۵۸۴۳ |
| ۸    | فیروزآباد  | ۹۰               | ۱۰۷         | ۹۸    |
| ۹    | قیروکارزین | ۲۹۷۴             | ۳۱۴۶        | ۲۵۸۸  |
| ۱۰   | کازرون     | ۸۱۴۸             | ۷۵۹۴        | ۸۸۰۱  |
| ۱۱   | لارستان    | ۱۱۲۵             | ۱۳۲۴        | ۱۰۶۱  |
| ۱۲   | لامرد      | ۱/۸              | ۱/۸         | ۲     |
| ۱۳   | مرودشت     | ۱/۸              | ۱/۶         | ۲     |
| ۱۴   | ممسنی      | ۶۲۴              | ۶۵۹         | ۶۳۵   |
| ۱۵   | مهر        | ۴/۲              | ۳/۹۸        | ۷     |
| ۱۶   | نیریز      | ۱۱۸/۶            | ۱۲۹/۷       | ۱۲۲   |

جدول ۸: مقایسه‌ی عمل کرد برآوردهای فی-هریوت  $\hat{Y}_{i,FH}$  و مستقیم،  $\hat{Y}_i$

| ردیف | شهرستان    | $\overline{MSE}(\hat{Y}_{i,FH})$ | $\overline{Var}(\hat{Y}_i)$ |
|------|------------|----------------------------------|-----------------------------|
| ۱    | استهبان    | ۲/۹۸                             | ۲/۷۳                        |
| ۲    | جهرم       | ۴۹۴/۳                            | ۵۶۶/۸                       |
| ۳    | داراب      | ۷۵۵/۰۱                           | ۱۵۱۳/۲                      |
| ۴    | زرین دشت   | ۰/۰۰۰۱                           | ۰/۰۰۰۲                      |
| ۵    | شیراز      | ۰/۰۰۰۲                           | ۰/۰۰۰۱                      |
| ۶    | فراشبند    | ۰/۰۰۰۱                           | ۰/۰۰۳                       |
| ۷    | فسا        | ۵۷۷/۵۳                           | ۱۱۸۷                        |
| ۸    | فیروزآباد  | ۰/۰۳۳                            | ۰/۲                         |
| ۹    | قیروکارزین | ۲۲/۸                             | ۴۹/۳۶                       |
| ۱۰   | کازرون     | ۴۵/۶                             | ۵۹/۷۲                       |
| ۱۱   | لارستان    | ۱۳/۵                             | ۲۵/۳۶                       |
| ۱۲   | لامرد      | ۰/۰۰۰۱                           | ۰/۰۰۰۱                      |
| ۱۳   | مرودشت     | ۰/۰۰۰۱                           | ۰/۰۰۰۱                      |
| ۱۴   | ممسنی      | ۱۳/۶                             | ۱۹/۱۵                       |
| ۱۵   | مهر        | ۰/۰۰۰۶                           | ۰/۰۰۰۴                      |
| ۱۶   | نیریز      | ۰/۰۱۷                            | ۰/۱                         |
|      | میانگین    | ۱۰۷/۸۶                           | ۲۱۳/۹۳                      |

## مراجع

- [1] Datta, G.S. and Lahiri, P. (2000). A Unified Measure of Uncertainty of Estimated Best Linear Unbiased Predictors in Small Area Estimation Problems, *Statistica Sinica*, **10**, 613-627.
- [2] Diggle, P.J. Heagerty, P. Liang, K. Y. and Zeger, S. L. (2002). *Analysis of longitudinal data, Second edition*. Oxford: Oxford University Press.
- [3] Fay, R.E. and Herriot, R.A. (1979). Estimation of Income from Small Places, An Application of James-Stein Procedures to Census Data, *Journal of the American Statistical Association*, **74**, 269-277.
- [4] Jiang, J. and Lahiri, P. (2006). Mixed model Prediction and Small Area Estimation, *Sociedad de Estadística e Investigación operativa*, **15**, No. 1.
- [5] Lohr, S.H. and Ybarra, L. (2008). Small Area Estimation when Auxiliary Information is Measured with Error, *Biometrika*, **95**, 4, 919-931.
- [6] Rao, J.N.K. (2003). *Small Area Estimation*, John Wiley Sons Inc, New York.
- [7] Searle, S.R. Casella, G. and McCulloch, C. E. (1992). *Variance components*, John Wiley Sons Inc, New York.