



کد گذاری

کمیته کد گذاری و طبقه‌بندی‌های آماری

آزمایش اول سرشماری عمومی نفوس و مسکن سال ۱۳۸۸



تعریف کدگذاری

- تبدیل اطلاعات کلیدی پرسشنامه‌های سرشماری از شکل داده‌های حرفی به داده‌های عددی (کد) به منظور استخراج نتایج بر اساس این کدها



رده های مورد نیاز کدگذاری

- کارشناس مسئول
- کدگذار
- بازبین کد



وظایف کارشناس مسئول کدگذاری

- شرکت در کلاس آموزشی و فراگیری شیوه‌های کدگذاری و به‌کارگیری نرم افزارهای مربوط
- هماهنگی با تحلیل‌گر سیستم برای تحویل و ارائه‌ی فایل‌های اطلاعاتی حوزه‌های سرشماری برای انجام کدگذاری
- کنترل و بررسی روزانه‌ی فعالیت‌ها و ارائه‌ی راهنمایی‌های لازم به کدگذاران و بازبینان کد
- نظارت بر به‌کارگیری نکات آموزشی راهنمای کدگذاری



وظایف کدگذار و بازین

- شرکت در کلاس آموزشی و فراگیری شیوه‌های کدگذاری و به‌کارگیری نرم‌افزارهای مربوط
- کدگذاری اقلام آماری تعیین‌شده در فرم‌های سرشماری به‌کارستن دقیق شیوه‌های کدگذاری برای تعیین کد مناسب
- هماهنگی با کارشناس مسئول کدگذاری و کمک گرفتن از وی هنگامی که تعیین کد مناسب با به‌کارگیری روش‌های گفته شده میسر نباشد.
- بازینی کدهای تعیین‌شده براساس راهنماهای مربوط و مجموعه‌ی کدها



طبقه‌بندی‌های مورد استفاده در کدگذاری

- طبقه‌بندی رشته‌ی تحصیلی
- طبقه‌بندی مشاغل
- طبقه‌بندی فعالیت‌های اقتصادی
- طبقه‌بندی استاندارد کشورها
- طبقه‌بندی‌های شهرستان و شهر



ستون‌هایی که باید کدگذاری شوند

الف) اطلاع مربوط به شهر و شهرستان:

پایین بخش پ در برگ نخست فرم^۲

۱. ستون ۹ جدول^۲(آدرس محل زندگی دیگر این فرد)
۲. سوال ۶، بخش ویژگی‌های عمومی افراد(آدرس محل مورد نظر)
۳. سوال ۲۲، بخش مهاجرت(آخرین محل زندگی فرد قبل از اقامت در این شهر یا آبادی)
۴. سوال ۲۷، بخش اینترنت، سواد و تحصیلات (محل تحصیل فرد)
۵. سوال ۳۹، بخش وضع فعالیت (شهر یا آبادی محل فعالیت فرد)



ستون‌هایی که باید کدگذاری شوند (ادامه)

ب) اطلاع مربوط به نام کشور

۱. سوال ۱۲، بخش ویژگی‌های عمومی افراد (محل تولد)
۲. سوال ۱۴، بخش ویژگی‌های عمومی افراد (تابعیت)
۳. سوال ۲۲، بخش مهاجرت (آخرین محل زندگی فرد قبل از اقامت در این شهر یا آبادی)



ستون‌هایی که باید کدگذاری شوند (ادامه)

ج) اطلاع رشته تحصیلی

- سوال ۳۱ بخش اینترنت، سواد و تحصیلات

د) اطلاع وضع فعالیت

- ۱. سوال ۳۵، بخش وضع فعالیت

ه) اطلاع شغل

- ۱. سوال ۳۷، بخش وضع فعالیت



شیوه‌های کدگذاری

- کدگذاری به کمک فهرست الفبایی
- کدگذاری با جستجوی واژه‌ها
- کدگذاری مستقیم



اطلاعات کمکی در کدگذاری

برای کدگذاری هر کدام از اطلاعات درج شده در پرسشنامه، برخی اطلاعات کمکی به شرح زیر در نرم افزار کدگذاری گنجانده شده و در دسترس کدگذار قرار می گیرد:

- برای کدگذاری رشته‌ی تحصیلی: اطلاعات مربوط به دوره، پایه یا مدرک تحصیلی، سن و وضعیت محصل یا غیر محصل بودن فرد



اطلاعات کمکی در کدگذاری (ادامه)

- برای کدگذاری شغل: اطلاعات مربوط به فعالیت اصلی محل کار و مدرک تحصیلی
- برای کدگذاری فعالیت اصلی محل کار: اطلاعات مربوط به شغل



تعداد نیروی برآورد شده و محقق شده

با توجه به متوسط کدهای هر پرسشنامه (حدود ۱۱ کد)، تعداد پرسشنامه ها (حدود ۲۶۰۰۰) و با احتساب ۱۰ درصد حجم بازبینی حدود ۳۰۰۰۰۰ کد پیش بینی شد که برای این حجم کار و با توجه به سوابق گذشته ۲۰ نفر کدگذار و بازبین پیش بینی شدند.

ولی با توجه به عدم تکمیل تعدادی از پرسشنامه ها در اجرای آزمایش، تغییر میزان بازبینی به ۵ درصد و انجام بخشی از کدگذاری به صورت خودکار این تعداد به ۷ نفر کاهش یافت (با احتساب کارشناسان مسئول)



حجم کار و مدت زمان

- کدگذاری: ۳۷۸ واحد کار ۱۰۰ تایی
- بازبینی: ۳۷۸ واحد کار ۵ تایی
- جمعاً: حدود ۴۰۰۰۰ کد (با احتساب بازگشت به کدگذاری)

کدگذاری از تاریخ ۲۴/۹/۸۸ تا تاریخ ۱۵/۱۰/۸۸ به مدت ۱۳ روز کاری به طول انجامید.
از این مدت ۱ روز به آموزش کدگذاری اختصاص داده شد.



موانع و مشکلات

- مشکلات اینترنتی
- مشکلات نرم افزار کدگذاری



مشکلات نرم افزار

- هنگام کدگذاری، حرکت به سمت پایین کادر واحد کار به صورت خودکار انجام نمی شود و کادر در محدوده بالایی ثابت می ماند.
- در صورت کدگذاری خودکار در ابتدای کار و کدگذاری دستی در مرحله بعد، روش کدگذاری الفبایی کارایی نخواهد داشت و پیشنهاد می شود که هنگام کلیک بر روی عبارات مورد نظر در هر واحد کار، آن عبارت به جای قرار گرفتن در باکس الفبایی در باکس جستجو قرار گیرد، تا نیازی به تایپ دوباره آن عبارت در باکس جستجو نباشد.



مشکلات نرم افزار

- از آنجاییکه رده بازبین و کدگذار به صورت جداگانه در نظر گرفته می شود شیوه بازبینی به شکلی طراحی شود تا امکان انجام بازبینی و نهایتاً تهیه گزارش مغایرت ها وجود داشته باشد.
- از آنجاییکه هنگام مواجه شدن با یک عبارت نامفهوم تا زمانی که کد مناسب برای آن تخصیص داده نشود واحد کار بسته نمی شود پیشنهاد می شود همانند نرم افزار سال ۱۳۸۵، کدی مانند (۲-) برای موارد نامشخص تعریف شود تا هم در بستن واحد کار و اتمام آن خللی ایجاد نشود و هم امکان تشخیص مواردی که واقعا به صورت نامشخص بودند و مواردی که کدگذار نتوانسته کدی به آن ها تخصیص دهد توسط کارشناس مسئول وجود داشته باشد.



مشکلات نرم افزار

- ترتیبی اتخاذ شود که پیش‌بینی گزارش‌گیری‌های مختلف از جمله تعداد و درصد مغایرت‌ها به تفکیک روش کدگذاری و روش جمع‌آوری اطلاعات برای هر طبقه‌بندی، تعداد و درصد موارد نامشخص به تفکیک روش کدگذاری و روش جمع‌آوری اطلاعات، تعداد و درصد کدگذاری خودکار به تفکیک روش جمع‌آوری اطلاعات امکان‌پذیر باشد.
- از آنجاییکه درصد موارد بازبینی شده جهت به‌دست آوردن خطای کدگذاری ضروری است درصد بازبینی با هماهنگی کمیته کدگذاری و طبقه‌بندی‌ها تعیین شود.



مشکلات نرم افزار

- حذف روش جستجوی مفهومی (content) تاثیر سوء در روند کدگذاری داشته است بطوریکه در بسیاری موارد کدگذاران مجبور بودند به کتاب کد مراجعه نمایند که به نظر می‌رسد این کار هنگامی که حجم کار بسیار بالا باشد مقدور نخواهد بود.
- بسیاری از مصادیقی که در کدگذاری خودکار شناسایی و کد نگرفته بودند عمدتاً ناشی از وجود فاصله بین حروف مصادیق بود در صورتیکه این مشکل نرم‌افزار حل شود درصد بیشتری از مصادیق در روش خودکار کد خواهند گرفت.
- هنگام جستجوی عبارات دو قسمتی در کادر جستجو مشکل وجود داشت بطوریکه هنگام جستجوی این عبارات، نرم افزار فقط مصادیقی را برای این موارد پیدا می‌کرد که در آن‌ها دو قسمت عبارت پشت سرهم بودند و مصادیقی که شامل هر دو قسمت عبارت بودند اما نه به صورت پشت سرهم ، توسط نرم افزار ارائه نمی‌شد.

کدگذاری خودکار

نوع طبقه‌بندی	تعداد کل کدهای درج شده	تعداد کدهای درج شده به صورت خودکار	تعداد کدهای درج شده توسط کدگذار	درصد کدهای درج شده به صورت خودکار
شهر	۲۷۳۱	۲۳۰۶	۴۲۵	۸۴٪
کشور	۵۷	۳۱	۲۶	۵۴٪
فعالیت	۱۸۴۷۱	۱۰۸۶	۱۷۳۸۵	۶٪
شغل	۱۳۱۷۱	۱۱۰۱	۱۲۰۷۰	۸٪
رشته تحصیلی	۱۵۹۴۱	۱۰۵۲۶	۵۴۱۵	۶۶٪
کل	۵۰۳۷۰	۱۵۰۴۴	۳۵۳۲۶	۳۰٪

موارد نامشخص

نوع طبقه‌بندی	تعداد موارد نامشخص	درصد موارد نامشخص
شهر	۱۳۴	۹/۴٪
کشور	۵	۸/۸٪
فعالیت	۲۶۸۸	۵/۱۴٪
شغل	۱۳۷	۱٪
رشته تحصیلی	۲۶۸	۷/۱٪
کل	۳۲۳۲	۴/۶٪

موارد نامشخص

اشکال در داده آمایی	اطلاعات ناقص (مامور یا پاسخگو)	اشکال کدگذاری یا طبقه‌بندی	علت نامشخص نوع طبقه بندی
%۶۴	%۳۰	%۶	شهر
%۷۳	%۶	%۲۱	رشته تحصیلی
%۱۰	%۵۳	%۳۷	فعالیت
%۲	%۶۲	%۳۶	شغل
%۵/۳۷	%۵/۳۷	%۲۵	کل



نتیجه گیری

در صورت رفع مشکلات نرم افزار کدگذاری و کندی اینترنت این روش برای آزمایش دوم و سرشماری ۹۰ مناسب می باشد ولی با شرایط موجود، به دلیل تشدید مشکلات اینترنت در استان ها کدگذاری به صورت غیر متمرکز امکان پذیر نمی باشد.