

Estudios sobre la Economía Española - 2020/25

**Series largas de algunos agregados económicos y demográficos
regionales: Actualización de RegData hasta 2019
(RegData y RegData Dem versión 6.0-2019)**

Angel de la Fuente
(FEDEA e Instituto de Análisis Económico (CSIC))

septiembre 2020

fedea

Las opiniones recogidas en este documento son las de sus autores y no coinciden necesariamente con las de FEDEA.

**Series largas de algunos agregados económicos y demográficos regionales:
Actualización de *RegData* hasta 2019
(*RegData* y *RegData Dem* versión 6.0-2019)**

Angel de la Fuente*
FEDEA e Instituto de Análisis Económico (CSIC)

Agosto de 2020

Palabras clave: renta, empleo y población regional, series homogéneas
Clasificación JEL/JEL Classification: E01, R1

Key words: regional income, employment and population, homogeneous series, Spain

1. Introducción

En esta nota se describe brevemente la última actualización de *RegData*, una base de datos que recoge los principales agregados económicos y demográficos de las regiones españolas durante las últimas seis décadas. En su mayoría, las series comienzan en 1950 o 1955 y se extienden hasta 2019. En años recientes, la información que ofrece *RegData* coincide con las últimas series oficiales que publica el INE. Estas series se enlazan con series oficiales anteriores y con otras fuentes con el fin de obtener series enlazadas homogéneas que se extienden hacia atrás tanto como ha sido posible.

La actualización de este año es algo más complicada que las anteriores pues se realiza tras la revisión estadística de 2019, que vendría a ser un cambio de base en la terminología tradicional. Por tanto, las nuevas series de Contabilidad Regional de España (CRE) del INE, que comienzan en 2016 (de aquí en adelante la CRE16 para abreviar) han de enlazarse con las series correspondientes de la versión anterior de REGDATA (versión 5.0). Esta complicación no afecta al módulo demográfico de REGDATA, que se extienden un año más por el procedimiento habitual de incorporar el nuevo dato del INE sin modificación alguna.

2. La revisión estadística de 2019

A finales de 2019, el INE publicó las primeras series de CRE que incorporan los resultados de la revisión estadística de 2019.¹ Como los antiguos cambios de base, esta revisión incluye una actualización de las fuentes estadísticas y métodos de estimación empleados en la elaboración de la CNE y la CRE e incorpora diversas mejoras metodológicas.² Como resultado de todo ello, las nuevas series de la CRE16 difieren de las anteriores (la CFRE10) en el año de referencia (2016 en este caso) y es necesario realizar un enlace entre ellas.

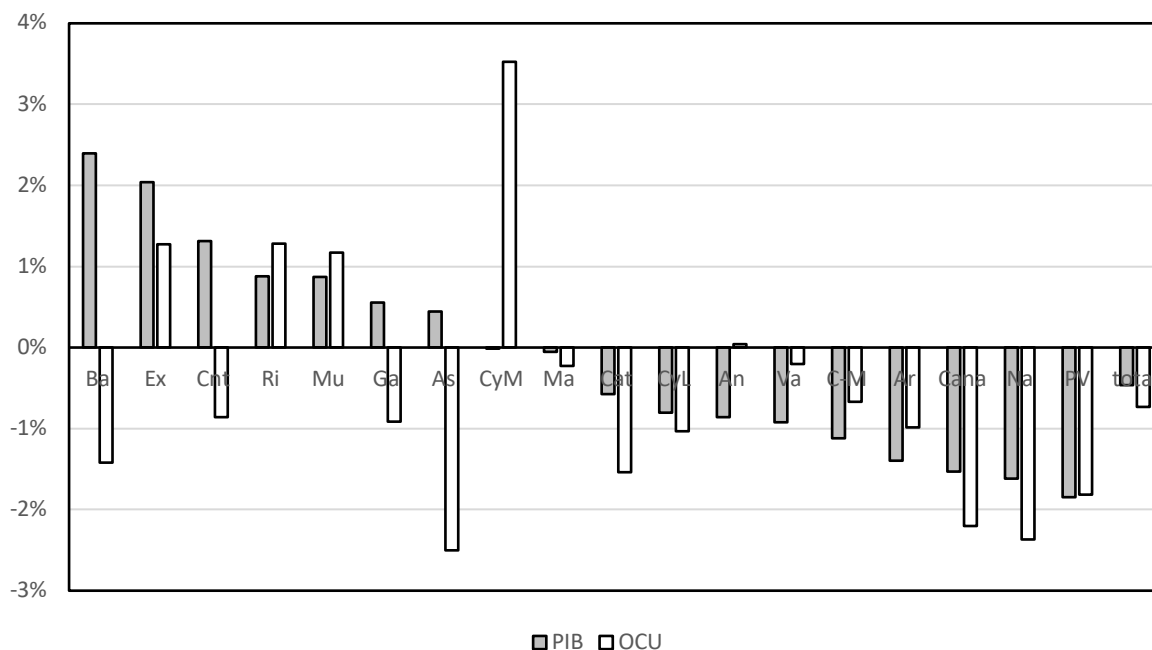
* Este trabajo forma parte de un proyecto de investigación cofinanciado por BBVA Research. Agradezco también la financiación del Ministerio de Economía, Industria y Competitividad bajo el proyecto ECO2017-87862-P.

¹ Véase https://www.ine.es/prensa/cre_2018_2.pdf

² Entre otras muchas novedades, se utilizarán datos de la inspección de la AEAT y de la Oficina de Lucha contra el Fraude para mejorar la estimación de la economía sumergida, se incorpora como fuente de información el directorio de cuentas de cotización de la Seguridad Social, y se utilizarán datos de caja para la medición de los ingresos por tributos y cotizaciones sociales. Para más detalles, véase INE (2019).

La revisión estadística de 2019 no afecta demasiado a los grandes agregados nacionales de output y empleo, pero sí tiene efectos significativos a nivel regional en muchos casos. Como se aprecia en el Gráfico 1, las nuevas estimaciones agregadas de PIB y ocupados nacionales en 2016 están en torno a medio punto porcentual por debajo de las estimaciones anteriores de la CRE10. A nivel regional, sin embargo, las diferencias son mayores, superando los dos puntos porcentuales en media docena de territorios. En algunos de ellos, además, el PIB y el empleo se revisan en direcciones contrarias.

Gráfico 1: Serie nueva CRE16 – serie CRE10, diferencia porcentual entre las dos últimas series de PIB y empleo de la CRE



- Clave: Ba: Baleares; Ex: Extremadura; Cnt: Cantabria; Ri: Rioja; Mu: Murcia; Ga: Galicia; As: Asturias; CyM: Ceuta y Melilla; Ma: Madrid; Cat: Cataluña; CyL: Castilla y León; An: Andalucía; Va: Comunidad Valenciana; C-M: Castilla-la Mancha; Ar: Aragón; Cna: Canarias; Na: Navarra; PV: País Vasco.

3. Procedimiento de enlace

El INE enlaza las nuevas series de la CRE16 con sus antecesoras para construir una “serie homogénea” 2000-18. El Instituto utiliza el procedimiento conocido como retropolación, consistente en extender hacia atrás las nuevas series utilizando las tasas de crecimiento de las antiguas. Manteniendo la metodología habitual en la construcción de REGDATA, en este trabajo el enlace entre la CRE16 y nuestras series anteriores se realizará utilizando el llamado *método mixto* desarrollado en de la Fuente (2014).

Esta metodología proporciona una vía intermedia entre los dos procedimientos más habituales en la literatura: la *interpolación* y la *retropolación*. El nombre del método hace referencia al hecho de que éste reparte a lo largo del tiempo la discrepancia entre las dos series que aflora en el momento del enlace de una forma que afecta tanto a las tasas de crecimiento de la serie más antigua como a su valor en el año base. Con este método, la senda de la corrección a la serie más antigua que ha de introducirse para obtener la serie enlazada se obtiene a partir del supuesto de que el "error de medición" que presenta la primera de estas series es proporcional a un componente "mal medido" del producto o del empleo total

cuya tasa bruta de crecimiento es un múltiplo constante de la tasa bruta de crecimiento del resto del agregado correspondiente.

Para describir con precisión el procedimiento, resulta necesario introducir algo de notación. Supongamos que en el momento T se produce alguna incidencia, tal como un cambio de base, que genera una discontinuidad en alguna variable de interés. Llamemos x_t a la serie original (no corregida) del agregado medida en logaritmos, y_t al logaritmo de la serie enlazada o corregida por la incidencia que queremos construir para $0 \leq t \leq T$ y d_t a la diferencia logarítmica entre ambas series. Interpretaremos d_t como el error de medición contenido en la serie original antes de la introducción del cambio metodológico e intentaremos adivinar su evolución en los años anteriores a T a partir del único valor conocido de esta magnitud -- el correspondiente al punto de enlace, d_T que se obtiene como la diferencia logarítmica entre las dos series a enlazar en ese punto. Con el método mixto, y_t se estima como

$$(1) \hat{y}_t = x_t + \hat{d}_t \quad \text{para } 0 \leq t \leq T \quad \text{con } \hat{d}_t = d_T \left(\frac{1}{2} \right)^{\frac{T-t}{H}}$$

donde H es la *vida media* del error de medición aflorado con el cambio metodológico, esto es, el tiempo necesario (yendo hacia atrás) para que el error observado en la serie más antigua en el momento T se reduzca a la mitad.

Para realizar el enlace, es necesario hacer algún supuesto sobre el valor del parámetro de vida media, H . Puesto que la vida media real no se conoce, es necesario hacer algún supuesto que será inevitablemente arbitrario, pero seguramente más razonable que los supuestos implícitos en los métodos tradicionales (que suponen que el error ha existido siempre hacia atrás en el caso de la retroprolación o que se origina en el año base de la serie antigua en el de la interpolación). Como se ha hecho habitualmente en la construcción de REGDATA cuando no existe información adicional que pueda ayudar a fijar ese parámetro, en lo que sigue se supondrá una vida media de 10 años, lo que implica un grado de persistencia de tal error que cabría considerar moderado

Cuando x e y son magnitudes monetarias, se plantea también la cuestión de cómo enlazar sus valores "reales" (en volumen o a precios constantes del año base). La opción más natural consiste en expresar ambas magnitudes en precios de un mismo año y aplicar después la ecuación (1) para construir el enlace. Un procedimiento equivalente que en la práctica resulta algo más sencillo consiste en enlazar por retroprolación los correspondientes índices de precios y , seguidamente, deflactar la serie nominal previamente enlazada utilizando esta serie de precios.

4. Construcción de algunas series auxiliares

Antes de proceder al enlace entre las dos fuentes, es necesario construir algunas variables que la CRE16 no proporciona directamente. Estas son, en particular, el número de puestos de trabajo y puestos de trabajo equivalentes a jornada completa (totales y asalariados) y las rentas totales del trabajo.

El número de puestos de trabajo equivalentes a jornada completa (totales y asalariados, *PTEJC* Y *PTASEJC*) en cada región se aproxima dividiendo las horas totales trabajadas (por el total de los ocupados y por los asalariados respectivamente) en la región de acuerdo con la CRE16 por la duración media estimada de la correspondiente jornada completa a nivel nacional. Esta variable se obtiene como el cociente entre el número de horas trabajadas y el número de puestos de trabajo equivalentes (totales y asalariados) en el conjunto de España utilizando datos de las nuevas series de la CNE y la Contabilidad Trimestral (CNTR) con base 2016 que, a diferencia de la CRE16, sí proporcionan información sobre estas variables. El cálculo es sólo aproximado porque supone que la duración media de la jornada completa es igual en todas las regiones, lo que no tiene por qué ser cierto.

La serie de puestos de trabajo (*PT*) se construye a partir de la serie de ocupados (*OCU*) de la nueva base de la CRE, de las series nacionales de ambas variables que ofrecen la CNE y la CNTR y de la información regional que ofrece la EPA sobre el porcentaje de los ocupados que declaran una segunda actividad además de su ocupación principal (*%SEGEMP*), elaborada a partir de los microdatos de la encuesta.

Calculamos en primer lugar el cociente entre el número de puestos de trabajo y el número de ocupados a nivel nacional con datos de la CNE16, obteniendo así un factor de conversión entre las dos variables a nivel agregado:

$$(2) RPTOC_t^{CNE} = \frac{PT_t^{CNE}}{OCU_t^{CNE}}$$

Seguidamente, dado que puestos de trabajo y ocupados cumplen la siguiente relación

$$(3) PT = (1 + \%SEGEMP) * OCU,$$

la información de la EPA sobre la fracción de los trabajadores que declaran una segunda ocupación, *%SEGEMP*, puede utilizarse para modular al alza o a la baja en cada región el factor de conversión nacional que nos da la CNE16. En particular, estimamos el número de puestos de trabajo de la región *r* en el año *t* como el producto

$$(4) \widehat{PT}_{rt}^{CRE} = OCU_{rt}^{CRE} * RPTOC_t^{CNE} * \frac{1 + \%SEGEMP_{rt}^{EPA}}{1 + \%SEGEMP_t^{EPA}}$$

donde la ausencia del subíndice regional, *r*, indica que estamos trabajando con el promedio nacional de la variable de interés.³

Para el caso del empleo asalariado, se procede exactamente de la misma forma y utilizando las mismas fuentes, incluyendo la información regional de la EPA sobre el porcentaje de los asalariados que declaran una segunda actividad además de su ocupación principal (*%SEGEMPAS*).

Finalmente, las rentas totales del trabajo (*RTL*) se construyen como sigue utilizando información de la CRE16. En primer lugar, se calcula el salario medio para el sector *s* en la región *r* dividiendo la remuneración total de los asalariados en el sector-región (*RAS*) por el número total de personas asalariadas (*AS*),

$$(5) w_{sr} = \frac{RAS_{sr}}{AS_{sr}}$$

³ El dato de ocupados correspondiente a la extra-regio se multiplica sólo por *RPTOC* puesto que la EPA no proporciona información sobre este territorio ficticio que permita afinar la estimación.

Seguidamente, el salario medio de cada sector-región se multiplica por el número total de ocupados (asalariados y no asalariados) correspondiente (*OCU*) y los resultados se agregan sobre sectores para obtener una estimación de las *rentas totales del trabajo (RTL)* en la región

$$(6) \text{RTL}_r = \sum_s w_{sr} \text{OCU}_{sr}$$

5. Algunos detalles del enlace

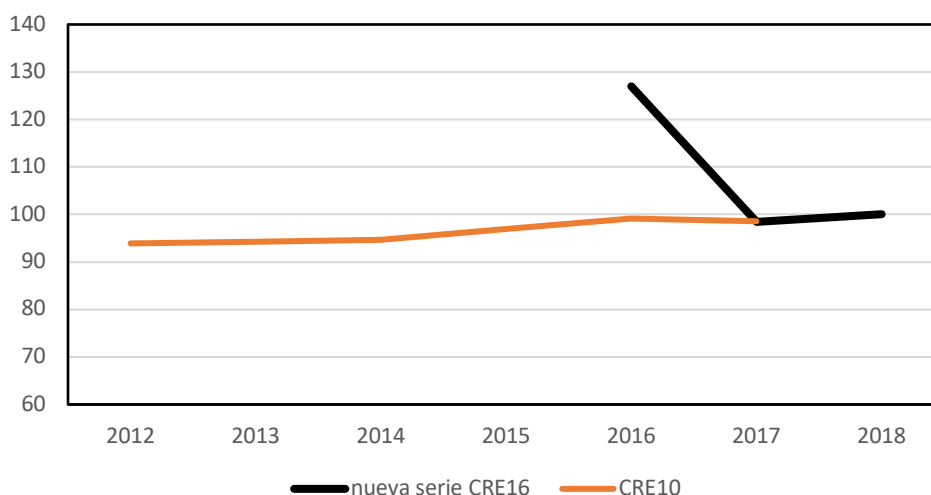
Para cada una de las variables de interés, el enlace entre las dos fuentes se realiza en primer lugar a nivel nacional por el procedimiento descrito en la sección 3. Seguidamente, se enlazan las correspondientes series regionales por el mismo procedimiento, tratando a la *extra regio* como un territorio más. Las series regionales preliminares obtenidas de esta forma se ajustan todas en la misma proporción de forma que su suma coincida exactamente con el correspondiente agregado nacional. Esta corrección se aplica a las variables de empleo, a la *RAS* y *RTL*, al *VAB* y al *PIB* a precios corrientes y constantes. Finalmente, las series definitivas de deflatores del *VAB* y el *PIB* se obtienen como el cociente entre las series definitivas de *VAB* (*PIB*) nominal y real. El procedimiento supone dar prioridad a las series nacionales construidas directamente sobre las que se obtendrían a partir de la suma de los enlaces de las variables regionales. Se ha optado por hacer así las cosas porque en algunos casos se dispone de mejor información a nivel nacional que regional.

Como se ha dicho más arriba, los deflatores del *VAB* y el *PIB* se enlazan por retropolación y las series enlazadas de deflatores se aplican a las correspondientes series enlazadas de producción a precios corrientes para obtener series a precios constantes (aproximados) de 2016. Puesto que la CRE ya no publica deflatores, estos se aproximan a partir de los correspondientes índices de volumen (que se renormalizan a 100 en 2016 antes de realizar los cálculos pertinentes).⁴ Para ello, se construyen en primer lugar series aproximadas de *PIB* "a precios constantes" de 2016 multiplicando los índices de volumen del *PIB* por el *PIB* del año base medido a precios corrientes. Los deflatores aproximados se calculan seguidamente como el cociente entre el *PIB* a precios corrientes y el *PIB* aproximado a precios constantes del año base. La serie de *PIB* a precios constantes obtenida a partir de esta serie preliminar de deflatores se ajusta en la misma proporción en todas las regiones de forma que su suma sobre comunidades autónomas (y la *extra-regio*) coincida con la serie nacional construida directamente por el mismo procedimiento. Finalmente, la serie final de deflatores se obtiene dividiendo la serie final de *PIB* a precios corrientes por la serie final de *PIB* a precios constantes. El procedimiento es el mismo en el caso del *VAB*.

Como se observa en el Gráfico 2, las nuevas series de horas trabajadas (totales y asalariadas) de Ceuta y Melilla presentan valores muy atípicos en su primer año de 2016. Para este año, la CRE16 da una cifra superior en un 30% a la CRE10, pero la diferencia entre las dos fuentes prácticamente desaparece en 2017 con la serie de la CRE16 revirtiendo al nivel de su antecesora. Para evitar que una observación atípica distorsione toda la serie enlazada, he eliminado el valor de 2016 de la nueva serie, sustituyéndolo por el que ofrece la serie anterior.

⁴ Rompiendo con la práctica habitual hasta el momento, los índices de volumen de la CRE16 se normalizan a 100 en 2015 en vez de en el año base de 2016.

Gráfico 2: Horas trabajadas en Ceuta y Melilla, CRE16 vs. CRE10



6. Variables disponibles

Los Cuadros 1 y 2 describen respectivamente las variables disponibles en el módulo central de *RegData* y en su módulo demográfico, *RegData Dem*. Los datos correspondientes se adjuntan en dos hojas de Excel. Una descripción detallada de la metodología y los datos originales de base que se han utilizado para construir las series de REGDATA pueden encontrarse en los trabajos citados en las referencias, que están disponibles en las webs de FEDEA (<http://www.fedea.net>) y de BBVA Research (<https://www.bbvaresearch.com/>).

El módulo central de *RegData* contiene datos de renta, precios, empleo y paro y rentas del trabajo. Se incluyen datos de PIB y VAB a precios corrientes y constantes de 2016 y a precios medios del período 1955-2019. El empleo total y asalariado se mide en personas, puestos de trabajo, horas y puestos de trabajo equivalentes a jornada completa. Las dos últimas variables comienzan en 1977 y 1981 respectivamente por falta de la información necesaria para construirlas en años anteriores. Finalmente, se ofrecen datos de remuneración de asalariados y salarios medios así como una serie de rentas totales del trabajo, que se aproximan imputando a cada trabajador no asalariado unas rentas salariales iguales a la remuneración media percibida por los asalariados de su misma región y sector.

En el módulo demográfico se ofrecen datos de población residente y su desglose por grupos de edad, nacimientos, defunciones y saldos migratorios. También se construye una serie de población residente corregida por migraciones no declaradas que intenta aproximar la población que realmente reside de forma habitual en cada territorio.

Cuadro 1: Contenido del módulo principal de *RegData*

| | | |
|------------------|--|-----------|
| VAB nom: | VAB nominal = valor añadido bruto a precios básicos, medido a precios corrientes (miles de euros corrientes) | 1955-2019 |
| VAB real: | VAB real = valor añadido bruto a precios básicos, medido a precios constantes (miles de euros de 2016) | 1955-2019 |
| Pvab10: | deflactor del VAB, año 2016 = 1 | 1955-2019 |
| Pvabmed: | deflactor del VAB a precios medios del período 1955-2019 | 1955-2019 |
| VAB pned | VAB real valorado a precios medios | 1955-2019 |

Cuadro 1: Contenido del módulo principal de RegData-- continuación

| | | |
|------------------|---|-----------|
| <i>PIB nom:</i> | PIB nominal = producto interior bruto a precios de mercado, medido a precios corrientes (miles de euros corrientes) | 1955-2019 |
| <i>PIB real:</i> | PIB real = producto interior bruto a precios de mercado, medido a precios constantes (miles de euros de 2016) | 1955-2019 |
| <i>Ppib10:</i> | deflactor del PIB, año 2016 = 1 | 1955-2019 |
| <i>Ppibmed:</i> | deflactor del PIB a precios medios del período 1955-2019 | 1955-2019 |
| <i>PIB pmed</i> | PIB real valorado a precios medios | 1955-2019 |
| <i>POBd:</i> | población residente o de derecho | 1955-2019 |
| <i>PARADOS:</i> | parados, miles de personas, promedio de los cuatro trimestres | 1955-2019 |
| <i>OCU:</i> | ocupados, miles de personas | 1955-2019 |
| <i>AS:</i> | asalariados, miles de personas | 1955-2018 |
| <i>PT:</i> | puestos de trabajo, en miles | 1955-2019 |
| <i>PTAS:</i> | puestos de trabajo asalariados, en miles | 1955-2018 |
| <i>H:</i> | horas efectivas anuales trabajadas por los ocupados, en millones | 1977-2018 |
| <i>HAS:</i> | horas efectivas anuales trabajadas por los asalariados, millones | 1977-2018 |
| <i>PTEJC:</i> | puestos de trabajo equivalentes a jornada completa, en miles | 1981-2018 |
| <i>PTASEJC:</i> | puestos de trabajo asalariados equivalentes a jornada completa, en miles | 1981-2018 |
| <i>RAS</i> | remuneración de asalariados, miles de euros corrientes | 1955-2018 |
| <i>RTL</i> | rentas totales del trabajo, miles de euros corrientes | 1955-2018 |
| <i>PARTL:</i> | participación corregida del trabajo en el VAB = RTL/VAB | 1955-2018 |
| <i>w:</i> | salario medio = $RAS/PTAS$ | 1955-2018 |
| <i>rml:</i> | remuneración media del trabajo = RTL/PT | 1955-2018 |

Cuadro 2: Contenido del módulo demográfico, RegData Dem

| | | |
|----------------------|--|-----------|
| <i>NAC(t)=</i> | nacimientos durante el año que comienza el 1 de julio de t | 1950-2018 |
| <i>DEF(t)=</i> | defunciones durante el año que comienza el 1 de julio de t | 1950-2018 |
| <i>SVEG(t)=</i> | saldo vegetativo del año que comienza el 1 de julio de $t = NAC(t) - DEF(t)$. | 1950-2018 |
| <i>POBd(t)=</i> | población de derecho o residente a 1 de julio de t. | 1950-2019 |
| <i>SMIG1(t)=</i> | saldo (in-)migratorio neto consistente con la serie de población de derecho = $POBd(t+1) - POBd(t) - SVEG(t)$ | 1950-2018 |
| <i>CORR_MIGND =</i> | Ajuste a la población de derecho por migraciones no declaradas | 1950-2019 |
| <i>POBRES_CORR =</i> | Población residente corregida por migraciones no declaradas = $POBd + CORR_MIGND$ | 1950-2019 |
| <i>SMIG2 =</i> | saldo (in-)migratorio neto consistente con la serie de población residente corregida = $POBRES_CORR(t+1) - POBRES_CORR(t) - SVEG(t)$ | 1950-2018 |
| <i>wPOBd_0-14 =</i> | peso de la población de 0 a 14 años en la población total de derecho | 1950-2019 |
| <i>wPOBd_65+ =</i> | peso de la población de 65 o más años en la población total de derecho | 1950-2019 |
| <i>wPOBd_0-15 =</i> | peso de la población de 0 a 15 años en la población total de derecho | 1950-2019 |
| <i>POBd_16+ =</i> | población de derecho de 16 o más años de edad | 1950-2019 |

Referencias

- de la Fuente, A. (2016a). "Series enlazadas de Contabilidad Regional para España, 1980-2014. (RegData_80-14, v4.3)." FEDEA, Estudios sobre economía española no. 2016-35, Madrid.
<http://www.fedea.net/datos-economia-regional-y-urbana/>
- de la Fuente, A. (2016b). "Series enlazadas de PIB y otros agregados de Contabilidad Nacional para España, 1955-2014. (RegDat_Nac versión 4.2)." FEDEA, Estudios sobre Economía Española no. 2016-02, Madrid.
<http://www.fedea.net/datos-economia-regional-y-urbana/>
- de la Fuente, A. (2016c). "Series largas de algunos agregados demográficos regionales, 1950-2015. RegData-Dem versión 5.0." FEDEA, Estudios de Economía Española no 2016-14, Madrid.
<http://www.fedea.net/datos-economia-regional-y-urbana/>
- de la Fuente, A. (2017). "Series enlazadas de algunos agregados económicos regionales, 1955-2014. Parte I: Metodología, VAB, PIB y puestos de trabajo (RegData_55-14, Versión 5.0-parte I)." FEDEA, Estudios sobre Economía Española no. 2017-08, Madrid.
<http://www.fedea.net/datos-economia-regional-y-urbana/>
- de la Fuente, A. (2017). "Series enlazadas de algunos agregados económicos regionales, 1955-2014. Parte II: Otras variables de empleo, rentas del trabajo y paro. RegDat Versión 5.0." FEDEA, Estudios sobre la Economía Española no. 2017-
<http://www.fedea.net/datos-economia-regional-y-urbana/>
- de la Fuente, A. (2019). "Series largas de algunos agregados económicos y demográficos regionales: Actualización de RegData hasta 2018. (RegData y RegData Dem versión 5.0-2018)." FEDEA, Estudios sobre la Economía Española no. 2019-19. Madrid.
<http://documentos.fedea.net/pubs/eee/eee2019-19.pdf>
- Instituto Nacional de Estadística (INE, 2019). "Revisión estadística 2019 de las operaciones de Contabilidad Nacional del INE. Proyecto técnico."
https://ine.es/normativa/leyes/cse/cambio_meto_Revisi%C3%B3nContabilidad2019.pdf
- Instituto Nacional de Estadística (INE, 2020a). Contabilidad Regional de España. Resultados. Serie contable 2016-19.
https://ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736167628&menu=resultados&idp=1254735576581#
- Instituto Nacional de Estadística (INE, 2020b). Contabilidad Nacional Anual de España. Principales agregados.
https://ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177057&menu=resultados&idp=1254735576581
- Instituto Nacional de Estadística (INE, 2020c). Contabilidad Nacional Trimestral de España. Principales agregados.
https://ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736164439&menu=resultados&idp=1254735576581
- Instituto Nacional de Estadística (INE, 2020d). Encuesta de población activa.
https://ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176918&menu=resultados&idp=1254735976595#!tabs-1254736195129
- Instituto Nacional de Estadística (INE, 2020e). Cifras de población.
https://ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176951&menu=resultados&idp=1254735572981

Véase también:

<https://www.bbvaresearch.com/category/tipos-de-documento/documento-de-trabajo/>