

GUÍAS DE PROFESORADO

ESTUDIOS ESTADÍSTICOS CON PERSPECTIVA DE GÉNERO

AUTORÍA
Lorenzo Escot Mangas



CÁTEDRA EXTRAORDINARIA
VALORES DEMOCRÁTICOS
Y GÉNERO



Instituto
Mujeres



 Instituto de
las Mujeres



instifem[®]

Catálogo de publicaciones de la Administración General del Estado

<https://cpage.mpr.gob.es>

Edita: Instituto de las Mujeres

Subdirección General de Estudios y Cooperación

Secretaría de Estado de Igualdad y para la

Eradicación de la Violencia contra las Mujeres

Autoría: Lorenzo Escot Mangas

enIPO: 050-25-049-X

El Instituto de las Mujeres no se hace responsable de las opiniones expresadas en esta guía.

1. ¿POR QUÉ INTRODUCIR LA PERSPECTIVA DE GÉNERO EN LA ESTADÍSTICA Y LA CIENCIA DE DATOS APLICADA?

La perspectiva de género es un enfoque analítico y práctico que busca identificar, cuestionar y transformar las desigualdades de género existentes en la sociedad. Cuando se habla de adoptar o incluir la perspectiva de género en el desarrollo de alguna actividad hablamos de identificar y eliminar de dicha actividad los posibles **estereotipos o roles sociales de género** que resultan en una discriminación de las mujeres por razón de su sexo biológico.

Los estereotipos de género son todas aquellas ideas, creencias, convenciones o prejuicios arraigados en la sociedad y que están relacionados con cuál es o cuál debería ser el rol que deberían desempeñar los hombres y las mujeres en nuestra sociedad (qué se espera de un varón y qué se espera de una mujer). Las personas utilizamos estos estereotipos como modelos para simplificar el mundo, para entender la realidad, para clasificar a otras personas en grupos con características o propiedades comunes. Los estereotipos de género atribuyen rasgos, actitudes, comportamientos y patrones a cada uno de los géneros. No son sólo características biológicas las que definen cada género sino, fundamentalmente, **construcciones sociales influenciadas por el contexto histórico y cultural**. Son creencias preconcebidas que se utilizan, en muchos casos de manera inconsciente y automática, como sustituto del razonamiento y análisis de los hechos objetivos. Los estereotipos son necesarios para el desarrollo normal de nuestras vidas. Los problemas surgen cuando esos estereotipos discriminan a un aparte de la sociedad, cuando impiden la igualdad de trato, la igualdad de oportunidades. Entonces esos estereotipos se convierten en dañinos y es necesario su erradicación (**Art. 1 de la Ley Orgánica 3/2007**).

Los estereotipos de género deben ponerse en cuestión cuando suponen una discriminación contra las mujeres, cuando sitúan a las mujeres en una situación de desigualdad de oportunidades frente a los varones. Por eso, adoptar la pers-

pectiva de género consiste, básicamente, en ser consciente de que, como seres sociales, también nosotros y nosotras podemos estar aplicando y reproduciendo unos estereotipos de género que discriminan a las mujeres, tal vez de manera inconsciente. Vivimos en sociedad y de una manera consciente o inconsciente somos partícipes de esos estereotipos.

Una vez identificada la presencia de estereotipos, la perspectiva de género implica también su erradicación de nuestra actividad, para evitar así perpetuar unos prejuicios no deseables. Los estereotipos de género adquieren fuerza en función de la cantidad de personas que los comparten. Así, cuantas más personas incluyan la perspectiva de género en su día a día, más fácil será corregirlos y caminar hacia el objetivo de la igualdad de oportunidades e igualdad de trato entre mujeres y hombres.

2. ESTUDIOS ESTADÍSTICOS CON PERSPECTIVA DE GÉNERO

Incorporar la perspectiva de género en la Estadística y la Ciencia de Datos aplicadas implica analizar los datos para detectar y estudiar la importancia de la asociación entre las normas o roles sociales y las diferencias o brechas de género. Incluir la perspectiva de género en los estudios estadísticos permite entender mejor las desigualdades y fomentar políticas más efectivas y justas. Los estudios estadísticos desempeñan un papel fundamental en el análisis de las diferencias sociales, culturales y económicas entre mujeres y hombres. Por un lado, es necesario proporcionar información veraz y objetiva de la magnitud de esas diferencias o brechas de género en los diferentes ámbitos de la vida (educación, empleo, salud, cultura, política, etc.). Y por otro, es imprescindible una adecuada diagnosis e interpretación de la evidencia que proporcionan los datos estadísticos.

· Datos desagregados por sexo:

Para cuantificar la magnitud y alcance las brechas de género es necesario disponer de datos fiables y veraces que refle-

jen la situación de varones y mujeres en los diferentes ámbitos de estudio. En este sentido, es necesario que las fuentes de datos (encuestas, paneles, registros administrativos, ...) que recopilen información referente a personas, incorporen al menos la variable sexo (sexo biológico). Sólo disponiendo de esa información, podrán analizarse los datos y presentar los resultados desglosados según sexo. La falta de desagregación por sexo frena el análisis de las brechas de género.

La ley Orgánica 3/2007 de 22 de marzo para la igualdad efectiva entre mujeres y hombres establece en su artículo 20 que, precisamente para poder analizar las brechas entre hombres y mujeres, las estadísticas, encuestas, y cualquier recolección de datos y estudios deben incluir sistemáticamente la variable sexo. A modo ilustrativo puede comprobarse cómo el INE utiliza la desagregación por sexo biológico (varones y mujeres) en todas sus estadísticas referidas a personas. De hecho, podríamos decir que esa desagregación por sexo es un mandato, al menos para los poderes públicos. Se reconoce así que la condición necesaria para introducir la perspectiva de género en las estadísticas es que toda la información estadística referente a personas debe ser recogida por sexo.

· **Sexo biológico, género, expresión de género y orientación sexual en los estudios estadísticos:**

Los términos sexo y género no son sinónimos, y no deberían utilizarse en las encuestas y estudios que recopilan datos de personas como si fueran conceptos equivalentes. Precisamente, cuando los estudios estadísticos diferencian sus datos según sexo y género de las personas, se suele utilizar **“varón” para referirse al sexo biológico masculino y “hombre” para referirse al género masculino** como construcción social. El sexo hace pues referencia al **sexo biológico**, es decir, a las características biológicas que diferencian a hombres y mujeres (cromosomas sexuales, genitales y hormonas que definen la fisionomía de las personas). La variable sexo es esencialmente binaria (varones y mujeres). La opción no binaria cuando se habla del sexo hace referencia a las personas intersexuales (entre los dos sexos), personas que nacen con características

sexuales biológicas que no se ajustan ni a las de hombre ni a la de mujer. Según la ACNUDH (Oficina del Alto Comisionado de las Naciones Unidas para el Desarrollo Humano) se calcula que hasta un 1,7 por ciento de la población mundial nace con características intersexuales.

Por su parte, el **género o la identidad de género**, se refiere a las identidades no tanto biológicas sino socioculturales y psicológicas de la sexualidad de las personas. La identidad de género es la percepción interna y subjetiva del género sexual que se atribuye cada persona a sí misma, independientemente del sexo biológico con el que haya nacido. Una persona puede ser **transgénero** (o transexual) cuando no se identifica con su sexo biológico o **cisgénero** (o cissexual) cuando no existe tal discrepancia. Muy asociado al concepto de identidad de género, pero diferente a éste, es el referente a la orientación sexual, que se refiere a la atracción emocional, física y sexual hacia otras personas. La **orientación sexual** de las personas no está definida ni por su sexo biológico ni tampoco por su identidad de género. Al preguntar por la orientación sexual se pueden utilizar diferentes categorías como heterosexual, gay, lesbiana, bisexual, queer, pansexual, asexual u otros.

La **identidad sexual** de una persona es por tanto multidimensional. Atañe al sexo biológico, la identidad de género y a la orientación sexual de la persona, pero también a la **expresión de género**, es decir, a la forma en que una persona expresa su identidad de género (por ejemplo, una expresión de género afeminada, masculina o androgina) independientemente de su sexo y su orientación sexual. La inclusión de la identidad de género, la orientación sexual y la expresión de género en los estudios estadísticos permite visibilizar la diversidad de la identidad sexual que va más allá de la opción binaria del sexo biológico (mujer o varón). Abrirse a otras opciones no binarias es reconocer que existen múltiples expresiones de la sexualidad humana, y son cruciales, por ejemplo, en las investigaciones sobre identidades sexuales que tratan de analizar las brechas, desigualdades, estereotipos y discriminaciones que sufre el colectivo LGTBIQ+.

No se necesita recoger información sobre las cuatro categorías (sexo, identidad de género, orientación sexual y expresión de género) en todas las estadísticas. Todo depende de qué dimensión se necesite para alcanzar el objetivo del estudio específico que se esté desarrollando. Por ejemplo, en el contexto de la Ley Orgánica 3/2007 de igualdad entre hombres y mujeres se establece que el objeto debe ser el sexo biológico y no el género o la orientación sexual. En los estudios en los que se amplía el análisis para considerar las desigualdades no sólo entre mujeres y hombres, sino también con personas de otras identidades de género, sí es necesario preguntar por la identidad sexual de forma más completa. Esto requiere especial cuidado a la hora de formular las preguntas, ya que se trata de cuestiones delicadas y sensibles sobre la intimidad de las personas, que en cierta manera podrían considerarse como una invasión de su privacidad. Por consiguiente, para reducir sesgos en las respuestas y errores de medida que afecten a la calidad del estudio, conviene explicar por qué se recogen esos datos, cómo se utilizarán y cómo se protegerán, considerando siempre el consentimiento informado y la no obligatoriedad de la respuesta a estos ítems. Una posible propuesta para conocer la identidad sexual de la persona participante en el estudio podría ser la siguiente:

- **Sexo:** ¿Cuál es su sexo (sexo biológico) ?: a) Varón; b) Mujer; c) Intersexual
- **Género/Identidad de género:** ¿Diría que su identidad de género actual coincide con el sexo que se le asignó al nacer?: a) Sí; b) No
- **Orientación sexual:** ¿Cuál es su orientación sexual?: a) heterosexual; b) gay; c) lesbiana; d) bisexual; e) queer; f) pansexual; g) asexual; f) demisexual; g) otros
- **Expresión de género:** ¿Cómo expresa su género en su actividad del día a día (en su forma de vestir, su apariencia física, su forma de hablar, de caminar, etc.)?: a) Expresión de género masculino; b) Femenino; c) Androgino; d) No binario; e) Fluido, f) Otros.

Es innegable que existe un estrecho vínculo entre sexo y género. Las desigualdades entre personas de diferente sexo responden y son el resultado de las desigualdades y estereotipos de género en las sociedades en las que convivan. Además, las brechas de género podrían referirse de manera más amplia a las diferencias entre mujeres, hombres y personas de otras identidades de género. A pesar de ello, cabe puntualizar que, en el ámbito de los estudios sobre brechas entre hombres y mujeres, el requerimiento es que la información estadística aparezca desagregada por sexo y no por género. El concepto de género va más allá de lo estrictamente biológico y hace alusión a identidades socioculturales, a percepciones internas y subjetivas de cada persona. El sexo es esencialmente una variable binaria (a excepción hecha de las personas intersexuales), y es la utilizada en los estudios sobre desigualdades entre hombres y mujeres. Por el contrario, la identidad de género, la orientación sexual y la expresión de género son otras dimensiones (no excluyentes sino complementarias al sexo biológico) de la identidad de sexo de una persona que incluyen toda una diversidad de opciones no binarias. La incorporación de estas otras dimensiones es más propia de las estadísticas y estudios que tratan de visibilizar y profundizar en las desigualdades, brechas, discriminaciones y estereotipos a los que se enfrenta el colectivo LGTBIQ+.

· **Indicadores de género:**

La incorporación sistemática del sexo en las fuentes estadística permite establecer indicadores que posibiliten un mejor conocimiento de las diferencias en los valores, roles, situaciones, condiciones, aspiraciones y necesidades de mujeres y hombres. Los indicadores de género son resúmenes de los datos estadísticos que desagregan la información según el sexo de las personas. Miden y visibilizan el alcance y la dimensión de las desigualdades o brechas en los diferentes ámbitos de la vida. Por ejemplo, los indicadores de brecha salarial media evidencian que los varones españoles en promedio ganan un salario un 20% mayor que el de las mujeres (INE, Encuesta de Estructura Salarial, 2022).

En términos estadísticos, los indicadores de género deben ayudar a detectar asociación entre la variable categórica sexo y las diferen-

tes variables que reflejan la vida de las personas (salarios, empleos, renta, pobreza, acceso a la sanidad, a la educación, a la cultura, a las ayudas sociales, a los puestos de responsabilidad tanto en el ámbito empresarial como en el de la política pública...). Suelen utilizarse tres grandes tipos de indicadores de género: los **índices de distribución**, que muestran la participación de las mujeres respecto al total de personas en una determinada categoría (por ejemplo participación de las mujeres entre el estudiantado de la universidad o en los puestos directivos de las grandes empresas); los **índices de segregación**, que permiten detectar asociación entre sexo y variables categóricas (como ocupaciones profesionales o titulaciones académicas) detectando aquellas categorías más segregadas por sexo; y por último los denominados indicadores de **brechas medias o medianas** para detectar asociación entre sexo y variables numéricas (como los ingresos, los salarios o la renta).

Índices de distribución: expresan la brecha entre el número de varones y de mujeres.

- **Índice de participación:** expresa el porcentaje de mujeres (varones) sobre el total de personas.
- **Índice de feminización o feminidad:** expresa la ratio o cociente entre mujeres y varones.

Índices de segregación: expresan la diferente participación de varones y de mujeres en las diferentes categorías de una variable cualitativa.

- **Índices de concentración:** expresan el porcentaje de mujeres (varones) en una determinada categoría en relación con el total de mujeres (varones) en el total de categorías.

- **Índices de representación:** mide la participación de las mujeres (frente al total de personas) en determinada categoría en relación (ratio) a la participación del total de mujeres en el conjunto de categorías.

- **Índices de disimilitud:** mide el porcentaje de población (mujeres o varones) que deberían cambiar de

categoría para alcanzar una representación equitativa de mujeres y varones en todas las categorías.

Brecha de género: expresa la diferencia entre varones y mujeres en la media o mediana de una variable numérica (normalmente esa diferencia se expresa en porcentaje respecto a la media o mediana de los varones).

La detección de **asociación estadística entre sexo y las diferentes variables socioeconómicas** puede interpretarse como evidencia de una posible situación de discriminación, directa o indirecta, contra el principio de igualdad de trato y de oportunidades entre mujeres y hombres. Para detectar esas situaciones de discriminación se han establecido una serie de umbrales de tolerancia a partir de los cuales se disparan las señales de alerta y se considera que las diferencias son lo suficientemente relevantes como para profundizar en sus potenciales causas. Se trata de determinar cuándo una brecha o diferencia en algún indicador de género es el reflejo de una posible situación de discriminación, de la existencia de barreras que impiden la igualdad de oportunidades entre hombres y mujeres. Estos **umbrales suelen establecerse en torno al 40-60% en los indicadores de participación, y del 25% en las ratios de feminización y representación** (por ejemplo, los consejos de administración de las empresas presentarán una infrarrepresentación de mujeres cuando el porcentaje de mujeres consejeras sea inferior al 40% del total de puestos en el consejo). **También el 25% es el umbral utilizado en las brechas medias y medianas.** De esta forma, cuando por ejemplo dentro de una misma empresa la brecha salarial en puestos de trabajo equivalentes supere el 25%, la empresa está obligada a informar de los motivos que justifican dicha brecha (Real Decreto 902/2020, de 13 de octubre, de igualdad retributiva entre mujeres y hombres).

• **Diferencias estructurales por razón del sexo en la modelización estadística:**

También la modelización multivariante permite detectar diferencias o brechas entre hombres y mujeres en diferentes ámbitos de interés. Para ello, tras la estimación o entrenamiento de un modelo

predictivo, pueden utilizarse los contrastes de cambio estructural (o multigrupo) para detectar diferencias según sexo: ¿son los parámetros estimados o las variables seleccionadas por nuestros algoritmos iguales para hombres y para mujeres? La detección de diferencias sistemáticas en los modelos predictivos estimados con las submuestras de varones y de mujeres puede interpretarse como el reflejo de la diferencia en el comportamiento, en las posibilidades de actuación, de trato o en la igualdad de oportunidades de hombres y mujeres.

A modo de ejemplo ilustrativo, la metodología de Oaxaca-Blinder permite descomponer las brechas salariales medias (o medianas) y detectar diferencias sistemáticas en las remuneraciones que reciben mujeres y hombres. Para ello se modelizan los salarios según las características observables de trabajadores y trabajadoras. Posteriormente se predice el salario que hubieran obtenido las mujeres, pero utilizando el modelo entrenado con la muestra de los varones. Esto permite dar respuesta a preguntas como ¿cuál sería el salario que en media hubiesen percibido las mujeres si sus características laborales y profesionales hubiesen sido tratadas/remuneradas como se hace con los hombres?

Tanto si utilizamos modelos de regresión más tradicionales como modelos predictivos que utilizan algoritmos más modernos de *machine learning*, *deep learning* o *inteligencia artificial*, cuando se utilizan datos referentes a personas en la modelización estadística resulta recomendable realizar estos contrastes de diferencias estructurales y reflexionar, en su caso, si esas diferencias reflejan una situación de discriminación, directa o indirecta, contra las mujeres.

3. ¿A QUÉ CUESTIONES DEBEMOS PRESTAR ESPECIAL ATENCIÓN?

Cuando se realizan estudios estadísticos existen una serie de problemas asociados a la consideración del género que pueden sesgar nuestras conclusiones y hacer que nuestros análisis carezcan de validez. Nos referimos a cuestiones referentes a la forma en la que recopilamos los datos, en la forma en que redactamos los cuestionarios para el trabajo de campo y en la propia interpretación de los resultados y elaboración de los

informes finales. Estos problemas pueden sesgar nuestras conclusiones y hacer que nuestros análisis carezcan de validez.

· **Evitar sesgos de selección muestral en la recolección de datos:**

El primero de los problemas que debemos intentar evitar en nuestros estudios se conoce como el **sesgo de selección muestral**, que es un error sistemático que ocurre cuando la muestra seleccionada para realizar un estudio y/o para estimar o entrenar un modelo, no es representativa del conjunto de la población que se pretende estudiar (o sobre la que se aplicará el modelo). Un ejemplo de este tipo de sesgos de selección muestral aparece cuando en la muestra de datos hay una sobrerepresentación de los varones frente a las mujeres, lo que puede derivar en conclusiones erróneas o sesgadas cuando se quieren generalizar al conjunto de personas (mujeres y hombres). Estos sesgos pueden ayudar a la perpetuación de las desigualdades o incluso agravarlas (por ejemplo, si sólo se consideran muestras de varones: en los estudios médicos sobre la efectividad de medicamentos o tratamientos; o en los estudios de mercado sobre preferencias de consumo; o en el entrenamiento de modelos de recomendación de películas o de obras literarias).

La mejor manera de evitar el sesgo de selección es mediante el **muestreo aleatorio**, donde cada persona tiene la misma probabilidad de ser seleccionada (independientemente de su sexo). Sin embargo, cuando este muestreo aleatorio no es posible, y con el fin de evitar los sesgos de selección muestral, habrá que asegurar la diversidad en la selección de la muestra de datos e identificar y controlar la utilización proporcional de las posibles variables que podrían estar influyendo en los resultados (las denominadas variables de confusión), como son el sexo, la edad, el nivel educativo, la etnia, el nivel socioeconómico, etc. Sólo así podremos asegurar la validez externa de nuestros estudios, es decir, que las muestras utilizadas son representativas del conjunto de la población objetivo de nuestros análisis.

· **Evitar el uso del lenguaje en los estudios estadísticos:**

Una fase crucial en todo estudio estadístico que implique fuentes primarias es el diseño del cuestionario y la redacción

de las preguntas que se realizarán a las personas entrevistadas. Incluir la perspectiva de género en el diseño y redacción del cuestionario requiere un enfoque autocrítico y cuidadoso que evite reproducir estereotipos de género. Y para ello resulta imprescindible el autodiagnóstico. Aunque en lo personal podemos identificarnos y sentirnos como personas plenamente igualitarias y feministas, vivimos inmersos en la sociedad y, por tanto, no somos ajenos a esos estereotipos que en muchas ocasiones actúan de una manera oculta e inconsciente. El objetivo último es realizar un diseño de la investigación lo más inclusiva e igualitaria posible con el objeto de evitar reproducir y perpetuar la estereotipación negativa de las mujeres.

Las preguntas mal diseñadas pueden influir en las respuestas de los encuestados. Si las preguntas están cargadas de estereotipos, pueden condicionar a las personas a responder de manera que refuercen esos prejuicios, incluso si no los comparten. Por ejemplo, preguntar “*¿Cómo se siente usted como mujer al conciliar su vida laboral con las responsabilidades domésticas?*” asume que las tareas domésticas son una responsabilidad de las mujeres (y no tanto de los varones). Las preguntas deben reflejar una comprensión inclusiva e igualitaria en términos de género. Si las preguntas reflejan y reproducen roles tradicionales de género (como la idea de que las mujeres deben ser las principales cuidadoras), se perpetúa una narrativa desigual. Por ejemplo, una pregunta como “*¿Qué actividades realiza como madre en el hogar?*” invisibiliza a los padres (varones) como corresponsables de las tareas de cuidado.

Por otra parte, reproducir estereotipos puede ser percibido como una falta de respeto hacia las personas encuestadas y podría comprometer la ética de la investigación y reducir las tasas de respuesta, especialmente entre mujeres que perciben la encuesta como insensible o prejuiciosa. Los datos recopilados mediante un cuestionario sesgado no solo serán menos útiles, sino que también pueden reforzar ideas problemáticas. Si el diseño de las preguntas no aborda las desigualdades estructurales o reproduce prejuicios, el análisis de los datos no será un reflejo real de las experiencias de las mujeres, sino una validación de estereotipos existentes. Esto afectará a nuestros estudios que

resultarán sesgados al no recoger la diversidad y complejidad de las experiencias humanas, especialmente las de las mujeres.

Otra recomendación práctica para evitar sesgos es la de revisar y analizar el lenguaje utilizado en la redacción de las preguntas. Revisar cómo redactamos y qué preguntamos. Evitando prejuicios sobre roles de género, y utilizando un lenguaje neutral e inclusivo. Por ejemplo, en lugar de preguntar “*¿ayuda usted a su pareja al cuidado de sus hijos?*”, mejor preguntar, de manera más neutral “*¿Cuántas horas dedica al cuidado de personas dependientes su cargo?*”. Es recomendable, por consiguiente, revisar algunas de las múltiples guías disponibles que nos ayudan a utilizar un lenguaje inclusivo y neutral (véase por ejemplo las guías publicadas por la Unidad de Igualdad de la Universidad Complutense de Madrid).

También en la diagnosis de los resultados y en los informes finales se debe tener un especial cuidado con la redacción, revisando el uso del lenguaje y evitando utilizar el género gramatical masculino como genérico (para nombrar a ambos sexos) cuando su uso resulte sexista, esto es, cuando invisibiliza de forma deliberada a las mujeres, cuando reproduce los estereotipos de género y reforza las desigualdades entre hombres y mujeres. Recordemos que la necesidad de incorporar la perspectiva de género en los estudios de Estadística y Ciencia de Datos responde a la necesidad de promover la igualdad, de eliminar prejuicios de género que discriminan a las mujeres. Aunque el uso del masculino genérico pretende abarcar a todas las personas, en la práctica existen muchas situaciones que no logra visibilizar explícitamente a quienes no se identifican con este género masculino. El argumento de que “así se ha hecho siempre” no es válido en un contexto de cambio social. El lenguaje está en constante evolución, reflejando las transformaciones culturales. Modificar el lenguaje incorporando formas más inclusivas no solo enriquece la comunicación, sino que también contribuye al cambio cultural y social necesario para alcanzar un mundo más justo e igualitario.



COORDINACIÓN

Isabel Tajahuerce Ángel

DISEÑO Y MAQUETACIÓN

Sonia Santandreu Ferragut