

## Géneros, Alcohol y Tabaquismo: Somos tan diferentes?

Daniel Donalisio - Felix Rojo Lapalma

8 de mayo de 2018

# Disclaimer

- El análisis fue realizado como anexo al Laboratorio 1 (bajo la sección Anexo - Laboratorio 2). Los archivos fuente pueden descargarse de [https://github.com/felixlapalma/diplodatos\\_2018](https://github.com/felixlapalma/diplodatos_2018)
- Hipótesis de uso: La presentación está orientada a generar un conjunto de información que pudiera ser aprovechada por organismos que realicen campañas de concientización relacionadas a los temas en estudio. Principalmente para contribuir a direccionar las acciones de un dado programa en forma y tiempo.
- Si bien las presentaciones tienden a ser autocontenidas, algunas partes requieren del hilo conductor que gestiona el presentador de la misma.

Terminología y notación: Usaremos en forma indistinta nombres en español o inglés para referirnos a género (gender), alcohol, tabaquismo (smoking) y cualquiera de las clases involucradas.

Hemos utilizado en forma indiscriminada tabaquismo (tabaco, etc) por *smoking*. Esto puede no ser correcto coloquialmente en cuanto el dataset no especifica el tipo de elemento consumido.

## Análisis

El presente abordaje esta enfocado en el análisis sobre igualdades y diferencias de género en cuanto a su actitud respecto al alcohol y el tabaquismo.

## Análisis

El presente abordaje esta enfocado en el análisis sobre igualdades y diferencias de género en cuanto a su actitud respecto al alcohol y el tabaquismo.

### Que pretendemos conocer y entender?

- La distribución de consumidores de tabaco por género.
- La distribución de consumidores de alcohol por género.
- La distribución etaria y por género de los que consumen alcohol y tabaco? Existe una edad en la que se da mas un hecho que otro?
- ¿Como es la distribución en el caso de los consumidores sociales, son mas proclives a una edad? ¿Se diferencian por género?

## Análisis

El presente abordaje esta enfocado en el análisis sobre igualdades y diferencias de género en cuanto a su actitud respecto al alcohol y el tabaquismo.

### Que pretendemos conocer y entender?

- La distribución de consumidores de tabaco por género.
- La distribución de consumidores de alcohol por género.
- La distribución etaria y por género de los que consumen alcohol y tabaco? Existe una edad en la que se da mas un hecho que otro?
- ¿Como es la distribución en el caso de los consumidores sociales, son mas proclives a una edad? ¿Se diferencian por género?

**La lista anterior no pretende ser exhaustiva ni extensiva, sino simplemente plantear un road-map que nos permita conocer un poco las igualdades y diferencias.**

## Acciones

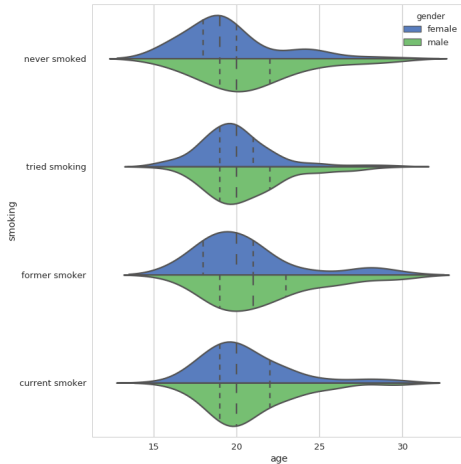
A partir del análisis, ser capaces de generar información útil que pudiera ser aprovechada por organismos que realicen campañas de concientización relacionadas a los temas en estudio. Principalmente a direccionar esas campañas en forma y tiempo. Es decir,

## Acciones

A partir del análisis, ser capaces de generar información útil que pudiera ser aprovechada por organismos que realicen campañas de concientización relacionadas a los temas en estudio. Principalmente a direccionar esas campañas en forma y tiempo. Es decir,

- **¿Tenemos que realizar un enfoque diferenciado respecto al género en el abordaje de temas como el alcoholismo y el tabaquismo?**

# Distribución

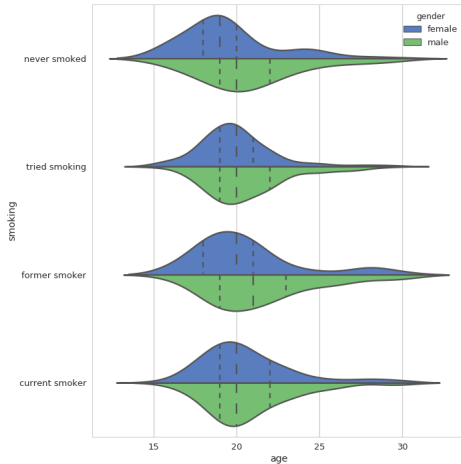


## never smoked

Existe una diferencia entre géneros, no en forma funcional, sino en su localización (observar los cuartiles en la misma figura). Esto podría indicar que es necesario un tratamiento diferenciado en cuanto al género.



# Distribución



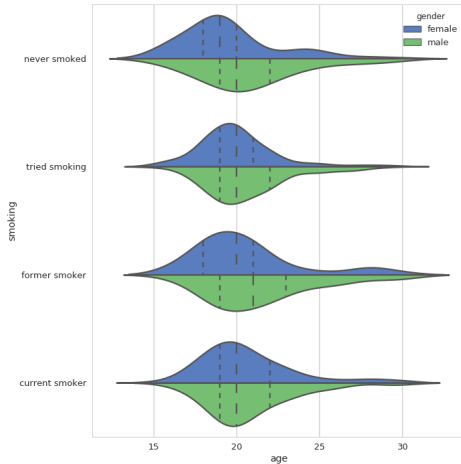
## never smoked

Existe una diferencia entre géneros, no en forma funcional, sino en su localización (observar los cuartiles en la misma figura). Esto podría indicar que es necesario un tratamiento diferenciado en cuanto al género.

## tried smoking

Muestra una paridad entre géneros respecto a su distribución (aunque el género masculino tiende a tener más intentos a mayor edad).

# Distribución



## never smoked

Existe una diferencia entre géneros, no en forma funcional, sino en su localización (observar los cuartiles en la misma figura). Esto podría indicar que es necesario un tratamiento diferenciado en cuanto al género.

## tried smoking

Muestra una paridad entre géneros respecto a su distribución (aunque el género masculino tiende a tener más intentos a mayor edad).

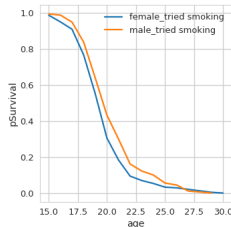
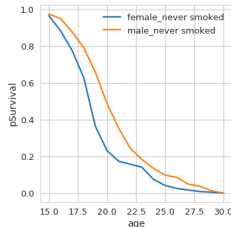
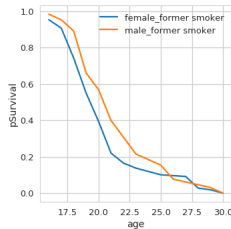
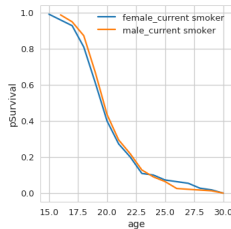
## former smoker

Similar a *never smoked*, la forma funcional del núcleo de la distribución parece estar desplazado pero no su forma.

## current smoker

No se observan diferencias evidentes para las/los actuales fumadores.

A los comentarios anteriores les podemos dar soporte extra considerando, dado un evento, su probabilidad de supervivencia, la cual no es mas que  $\Phi(t) = 1 - FDA(t)$  (para FDA función de densidad acumulada).

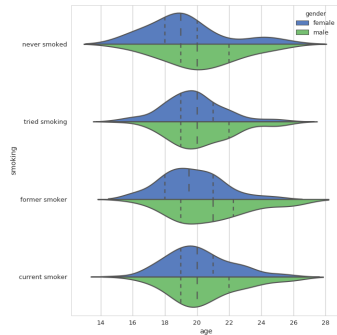
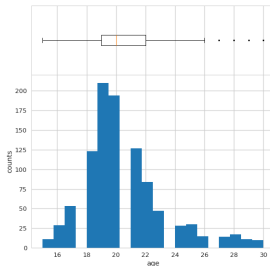


## Supervivencia de un Comportamiento

Las probabilidades de supervivencia graficadas vemos que hacen plausibles las suposiciones en cuanto a diferencias o similitudes (en las distribuciones). En virtud de estas podríamos generar ciertos indicativos, en este caso que, el enfoque respecto a los grupos de consumo debería empezar en forma mas temprana para el género femenino.

# Outliers

El analisis de outliers lo realizamos sobre la distribucion de edades. Algunos comentarios en relacion al impacto sobre los generos, se pueden observar en el codigo utilizado. Igualmente en las figuras siguientes mostramos las distribuciones resultantes.



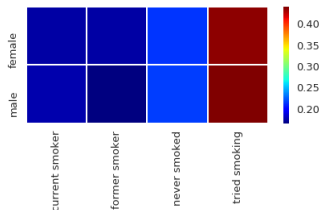
A partir de las distribuciones graficadas observamos que las mismas conservan esencialmente la misma forma (no se observan mayores corrimientos ni cambios en la modalidad de las distribuciones). En la presente situación corresponden al tabaquismo (más adelante mostramos las mismas para el alcoholismo).

Veamos (ya sin la distribución de edad) como se distribuye por género el tabaquismo,

	current smoker	former smoker	never smoked	tried smoking
female	70	70	86	176
male	48	45	59	120

Un primer vistazo a la tabla no arroja grandes diferencias por clase. Y se observa en forma similar para las cuentas normalizadas,

	current smoker	former smoker	never smoked	tried smoking
female	0.174129	0.174129	0.213930	0.437811
male	0.176471	0.165441	0.216912	0.441176



En este formato vemos que son similares las proporciones por clase (tanto numericamente como graficamente). Veamos que nos puede decir el test  $\chi^2$ .

Es decir buscamos responder si las diferencias observadas en los grupos son relacionadas al azar o tienen una correlacion o de otra forma *si la actitud respecto al tabaquismo se distribuye de la misma manera independiente del género.*

	$\chi^2$	$p - value$	$dof$
valor	0.08	0.99	3

Es decir buscamos responder si las diferencias observadas en los grupos son relacionadas al azar o tienen una correlacion o de otra forma *si la actitud respecto al tabaquismo se distribuye de la misma manera independiente del género.*

	$\chi^2$	$p - value$	$dof$
valor	0.08	0.99	3

### Recordemos...

Cuanto mayor sea el valor de  $\chi^2$ , menos verosímil es que la hipótesis nula (que asume la igualdad entre ambas distribuciones) sea correcta. De la misma forma, cuanto más se aproxima a cero el valor de  $\chi^2$ , más ajustadas están ambas distribuciones.

Es decir buscamos responder si las diferencias observadas en los grupos son relacionadas al azar o tienen una correlacion o de otra forma *si la actitud respecto al tabaquismo se distribuye de la misma manera independiente del género.*

	$\chi^2$	$p - value$	dof
valor	0.08	0.99	3

### Recordemos...

Cuanto mayor sea el valor de  $\chi^2$ , menos verosímil es que la hipótesis nula (que asume la igualdad entre ambas distribuciones) sea correcta. De la misma forma, cuanto más se aproxima a cero el valor de  $\chi^2$ , más ajustadas están ambas distribuciones.

En función de los resultados podemos decir que las distribuciones parecen tener un origen común (es decir la actitud respecto al tabaquismo parece tener un mismo comportamiento entre géneros). La significancia estadística la vemos en terminos del  $p$ -valor. El valor  $p$  nos muestra la probabilidad de haber obtenido el resultado que hemos obtenido si suponemos que la hipótesis nula es cierta. En nuestro caso  $p > 0,99$  con lo cual **NO** podemos rechazar la hipotesis de que los datos provengan de una misma distribución y las diferencias solo sean debidas al azar.



En relación al procesamiento de los datos, podemos realizar un par de observaciones,

En relación al procesamiento de los datos, podemos realizar un par de observaciones,

### Sobre las distribuciones, y comportamientos condicionados por edad

El análisis realizado indica que el comportamiento en relación a la actitud a ciertas fases del tabaquismo, parece diferenciar los generos. Particularmente la fases de *nunca consumió* (*never smoked*) e *intentó consumir* (*tried smoking*) parecen exigir un tratamiento por género.

En relación al procesamiento de los datos, podemos realizar un par de observaciones,

### Sobre las distribuciones, y comportamientos condicionados por edad

El análisis realizado indica que el comportamiento en relación a la actitud a ciertas fases del tabaquismo, parece diferenciar los generos. Particularmente la fases de *nunca consumió* (*never smoked*) e *intentó consumir* (*tried smoking*) parecen exigir un tratamiento por género.

### Sobre el comportamiento en general

El test  $\chi^2$  parece indicar un origen comun en el comportamiento general o actitud hacia el consumo de tabaco.

En relación al procesamiento de los datos, podemos realizar un par de observaciones,

### Sobre las distribuciones, y comportamientos condicionados por edad

El análisis realizado indica que el comportamiento en relación a la actitud a ciertas fases del tabaquismo, parece diferenciar los generos. Particularmente la fases de *nunca consumió* (*never smoked*) e *intentó consumir* (*tried smoking*) parecen exigir un tratamiento por género.

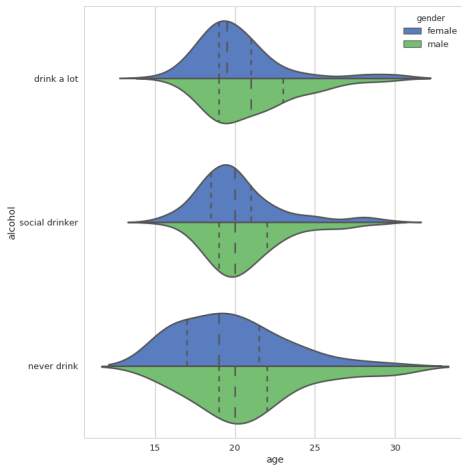
### Sobre el comportamiento en general

El test  $\chi^2$  parece indicar un origen comun en el comportamiento general o actitud hacia el consumo de tabaco.

### Acciones

Las comentarios previos indican que **No** somos tan diferentes - por género - en cuanto a nuestra actitud hacia el tabaco. Es decir un enfoque unificado seria de utilidad para ambos géneros (en consideración). Al margen que deberian ajustarse ciertos tiempos de aplicación de estos enfoques.

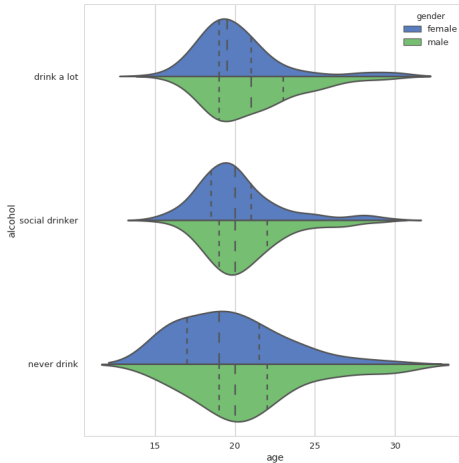
# Distribución



## never drink

Parece existir una diferencia entre géneros, no en forma funcional para la distribución sino en su localización (observar cuartiles en la misma figura).

# Distribución



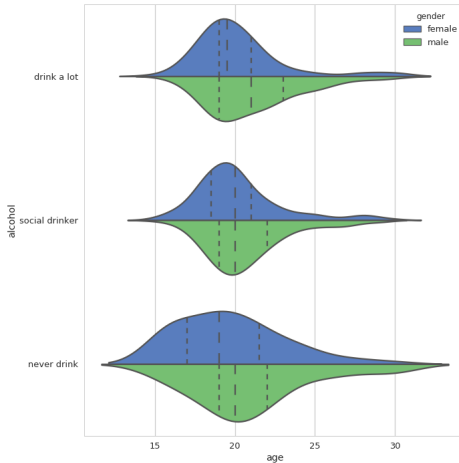
## never drink

Parece existir una diferencia entre géneros, no en forma funcional para la distribución sino en su localización (observar quartiles en la misma figura).

## social drinker

Muestra una paridad entre géneros respecto a su distribución (aunque el género masculino tiende a tener mas intentos a mayor edad).

# Distribución



## never drink

Parece existir una diferencia entre géneros, no en forma funcional para la distribución sino en su localización (observar cuartiles en la misma figura).

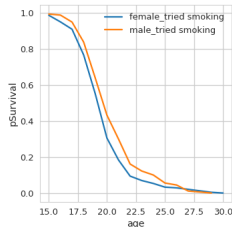
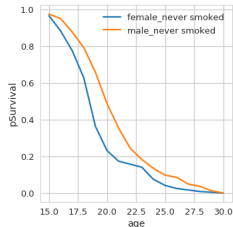
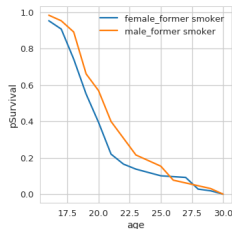
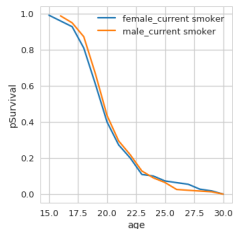
## social drinker

Muestra una paridad entre géneros respecto a su distribución (aunque el género masculino tiende a tener mas intentos a mayor edad).

## drink a lot

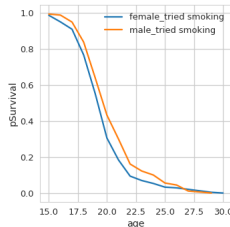
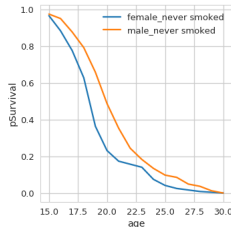
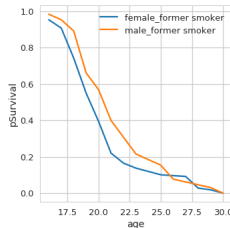
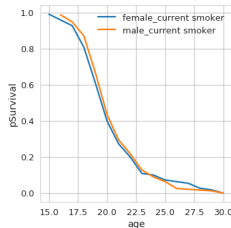
Muestra una paridad entre géneros (aunque la mediana para el género masculino se encuentra en una edad mayor que en el caso femenino).

A los comentarios anteriores les podemos dar soporte extra considerando, dado un evento, su probabilidad de supervivencia, la cual no es mas que  $\Phi(t) = 1 - FDA(t)$  (para FDA función de densidad acumulada).





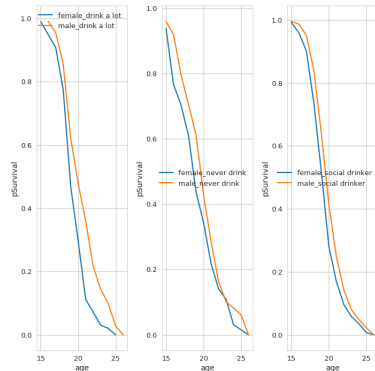
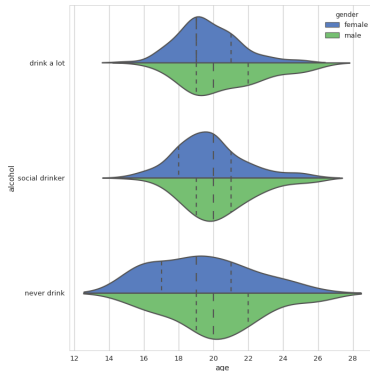
A los comentarios anteriores les podemos dar soporte extra considerando, dado un evento, su probabilidad de supervivencia, la cual no es mas que  $\Phi(t) = 1 - FDA(t)$  (para FDA función de densidad acumulada).



### Supervivencia de un comportamiento

Las probabilidades de supervivencia graficadas vemos que dan soporte a las suposiciones en cuanto a diferencias o similitudes (en las distribuciones). En virtud de estas podríamos generar ciertos indicativos, en este caso, están indicando que se debería abordar en forma diferenciada el consumo excesivo de alcohol en el género masculino.

# Outliers



La remoción de outliers no parece alterar en forma significativa ninguna de las distribuciones representadas. Esto con la salvedad que en el caso de grandes consumidores se marca mas definidamente la existencia de cierta asimetría en la permanencia de consumidores a edades mayores en el género masculino sobre el género femenino.

Veamos (sin condicionar por edades) como se distribuye por género el consumo de alcohol,

	drink a lot	never drink	social drinker
female	74	40	288
male	79	33	160

Un primer vistazo a la tabla arroja cierta asimetría en una de las clasificaciones. Y si observamos en cuentas normalizadas,

	drink a lot	never drink	social drinker
female	0.184080	0.099502	0.716418
male	0.290441	0.121324	0.588235



En este formato vemos que las proporciones por clase (tanto numericamente como graficamente) tienden a diferenciarse. Veamos que nos puede decir el test  $\chi^2$  de Person respecto a esto.

$\chi^2$ 

Es decir buscamos responder si las diferencias observadas en los grupos son relacionadas al azar o tienen una correlación o de otra forma *si la actitud respecto al alcohol se distribuye de la misma manera independiente del género.*

	$\chi^2$	$p - value$	$dof$
valor	12.8	0.0016	2

Es decir buscamos responder si las diferencias observadas en los grupos son relacionadas al azar o tienen una correlación o de otra forma *si la actitud respecto al alcohol se distribuye de la misma manera independiente del género.*

	$\chi^2$	$p - value$	$dof$
valor	12.8	0.0016	2

En función de lo obtenido podemos decir que las distribuciones parecen tener un origen distinto (es decir la actitud respecto al consumo de alcohol parece tener un comportamiento diferente entre géneros). La significancia estadística la vemos en términos del  $p$ -valor. El valor  $p$  nos muestra la probabilidad de haber obtenido el resultado que hemos obtenido si suponemos que la hipótesis nula es cierta. En nuestro caso  $p < 0,001$  ( $p\% < 0,16\%$ ) con lo cual **SI** podemos rechazar la hipótesis de que los datos provengan de una misma distribución.

En relación al procesamiento de los datos, podemos realizar un par de observaciones,

En relación al procesamiento de los datos, podemos realizar un par de observaciones,

### Sobre las distribuciones, y comportamientos condicionados por edad

El análisis realizado indica que el comportamiento en relación a la actitud a ciertas fases del alcoholismo, parece diferenciar los generos. Particularmente dos de las tres fases indican un comportamiento diferenciado en cuanto a lo temporal. Por caso, el genero femenino parece tener una transición mas acelerada en el caso de alto consumo, lo cual en si mismo parece mejor, puesto que lo abandona en forma mas temprana, pero seria necesario entender esta situación. En el lado opuesto, el género masculino en terminos de edad alcanza mas tarde esta condicion y su permanencia es mas extendida, con lo que exigiria un analisis del porque de su permanencia para orientar acciones en este punto.

En relación al procesamiento de los datos, podemos realizar un par de observaciones,

### Sobre las distribuciones, y comportamientos condicionados por edad

El análisis realizado indica que el comportamiento en relación a la actitud a ciertas fases del alcoholismo, parece diferenciar los generos. Particularmente dos de las tres fases indican un comportamiento diferenciado en cuanto a lo temporal. Por caso, el genero femenino parece tener una transición mas acelerada en el caso de alto consumo, lo cual en si mismo parece mejor, puesto que lo abandona en forma mas temprana, pero seria necesario entender esta situación. En el lado opuesto, el género masculino en terminos de edad alcanza mas tarde esta condicion y su permanencia es mas extendida, con lo que exigiria un analisis del porque de su permanencia para orientar acciones en este punto.

### Sobre el comportamiento en general

El test  $\chi^2$  parece indicar un origen diferenciado en el comportamiento general o actitud hacia el consumo de alcohol.



## Sobre las distribuciones, y comportamientos condicionados por edad

## Sobre el comportamiento en general

## Acciones

Las comentarios previos indican que **Si** parecemos mantener diferencias - por género - en cuanto a nuestra actitud hacia el alcohol. Es decir un enfoque diferenciado seria de utilidad para ambos géneros (en consideración). Asimismo en este enfoque diferenciado deberian ajustarse ciertos tiempos de aplicacion (consistentes para cada genero).

En esta sección recolectamos simplemente el conjunto de acciones y las dejamos en consideración,

En esta sección recolectamos simplemente el conjunto de acciones y las dejamos en consideración,

### Acciones

Las comentarios previos indican que **No** somos tan diferentes - por género - en cuanto a nuestra actitud hacia el tabaco. Es decir un enfoque unificado seria de utilidad para ambos géneros (en consideración). Al margen que deberian ajustarse ciertos tiempos de aplicacion de estos enfoques.

En esta sección recolectamos simplemente el conjunto de acciones y las dejamos en consideración,

### Acciones

Los comentarios previos indican que **No** somos tan diferentes - por género - en cuanto a nuestra actitud hacia el tabaco. Es decir un enfoque unificado sería de utilidad para ambos géneros (en consideración). Al margen que deberían ajustarse ciertos tiempos de aplicación de estos enfoques.

### Acciones

Los comentarios previos indican que **Si** parecemos mantener diferencias - por género - en cuanto a nuestra actitud hacia el alcohol. Es decir un enfoque diferenciado sería de utilidad para ambos géneros (en consideración). Asimismo en este enfoque diferenciado deberían ajustarse ciertos tiempos de aplicación (consistentes para cada género).

# Preguntas?

## Gracias!