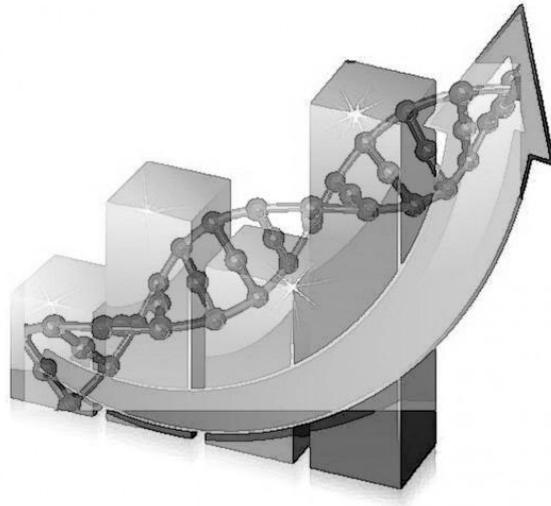


ESTADÍSTICA EN LA GENÉTICA



Equipo:

María Carcelén González

Andrés López Barba

David Martínez González

Marta Selva Castelló

Categoría: Secundaria

Tutor: Emilio Sánchez Alcantud

ESTADÍSTICA EN LA GENÉTICA

Índice:

1. Introducción	pag. 2
2. Objetivos	pag. 2
3. Características del estudio	pag. 3
4. Estudios realizados y resultados	pag. 6
5. Conclusiones	pag. 23
6. Posibles mejoras y extensiones del estudio	pag. 25
7. Anexos	pag. 26

1. Resumen

Al principio no sabíamos sobre qué escribir. Durante mucho tiempo reflexionamos buscando algún tema que pudiera interesar a la gente y fuera útil. Dado que habíamos estudiado un poco de genética en la asignatura de Biología y Geología y que nuestros compañeros se mostraban interesados en comprender el tema, decidimos que sobre ello trataría nuestro trabajo de estadística. Hemos estudiado la incidencia en las personas nacidas en Albacete de determinados fenotipos. Dado que hemos necesitado utilizar algunas técnicas (como el uso de tablas de contingencia) que no habíamos estudiado hasta el momento, nuestro profesor decidió adelantar algunos contenidos imprescindibles para que pudiéramos desarrollar nuestro trabajo.

Este trabajo ha sido posible gracias a la colaboración de cientos de personas que amablemente decidieron dedicarnos unos minutos para realizar la encuesta. Sin la ayuda de nuestros profesores de Matemáticas y de Biología el estudio nunca habría sido posible.

2. Objetivos

El objetivo principal fue el de conocer la incidencia en las personas nacidas en Albacete de algunos de los caracteres hereditarios de la especie humana que se pueden observar fácilmente.

Objetivos específicos:

1. Describir la incidencia de cada fenotipo en la muestra.
2. Estudiar si hay relación entre ser zurdo o no y cruzar o no piernas, brazos o dedos de una determinada forma.
3. Estudiar si hay relación entre el color de los ojos y del pelo.
4. Contrastar la proporción de personas con RH negativo en Albacete con la proporción nacional.
5. Estudiar si hay diferencias entre hombres y mujeres para cada uno de los fenotipos en estudio.
6. Estudiar si ha habido cambios a lo largo de los años en cada uno de los fenotipos analizados.

3. Características del trabajo

Para seleccionar la muestra tomamos los datos del Padrón en Albacete a 01 de enero de 2016 (Fuente: Instituto nacional de Estadística). De ahí obtuvimos los totales de personas empadronadas en Albacete por sexo e intervalos de edad de amplitud 10 años. Sobre los totales y usando la siguiente fórmula
$$\left(\frac{N \cdot z_{\alpha}^2 \cdot p \cdot q}{\text{precisión}^2 \cdot (N-1) + z_{\alpha}^2 \cdot p \cdot q}\right)$$
, obtuvimos el total de personas a las que había que encuestar. Nos salía una muestra de 2362 elementos. Dado su gran tamaño, lo redujimos al 10% (236 elementos como mínimo) y de manera proporcional al número total de personas por sexo y grupo de edad (muestreo estratificado). De esta forma obtuvimos el tamaño muestral correspondiente en cada caso. Nos repartimos por barrios para abarcar la mayor parte de la ciudad. Dado que no todas las personas censadas han nacido en Albacete (al decir Albacete nos referimos también a la provincia) o son hijos de personas nacidas en Albacete, al ir a entrevistar a cada persona lo primero que hacíamos era preguntarle si era natural de Albacete o lo era alguno de sus padres. Si la respuesta era negativa la descartábamos, agradeciéndole su amabilidad al respondernos, y seguíamos buscando. Así lo hicimos hasta completar la muestra. En el anexo hemos puesto la encuesta que realizamos.

Para tratar los datos hemos utilizado el software para Estadística R y en concreto el paquete RCommander. Después de recoger los datos, numeramos cada encuesta para tener más fácil la identificación de cada persona, por si necesitábamos recurrir a ella para aclarar alguna cuestión. En el archivo que contiene los datos hay una variable que indica el número de caso que corresponde a cada persona encuestada. Ese mismo número aparece en la encuesta de la persona correspondiente. A la hora de pasar los datos, lo hicimos primero en Excel y después, desde RCommander, importamos los datos. A partir de ahí, ya hicimos todo el trabajo con R. Como íbamos a aplicar técnicas estadísticas que nosotros desconocíamos, el profesor nos preparó dos presentaciones en Powerpoint sobre tablas de contingencia y el test de la Chi-cuadrado y sobre el contraste para la proporción y la diferencia de proporciones. Nos explicó en horario fuera de clase, los conceptos y procedimientos básicos para que lo pudiéramos entender y trabajar con ello. Desconocemos la fundamentación matemática, pero sí que conocemos un poco cómo funcionan estos tipos de test y sabemos interpretar los resultados que nos da el programa R.

Los fenotipos objeto de estudio han sido los siguientes:

1. Posibilidad de sacar la lengua enrollada
2. Lóbulo de la oreja pegado a la piel
3. Presencia de pelos en la segunda falange de los dedos de la mano
4. Ser zurdo
5. Cruzar naturalmente el brazo derecho sobre el izquierdo
6. Cruzar naturalmente la pierna izquierda sobre la derecha
7. Cruzar las manos con el dedo índice de la mano izquierda sobre el índice de la derecha
8. Dedo índice de menor longitud que el dedo anular.
9. Color de pelo claro
10. Ojos claros
11. RH negativo

Las variables asociadas a estos fenotipos son cualitativas, ya que sólo caben dos posibles respuestas (Sí o NO) y no son números

También hemos usado variables para agrupar a la población. Son las siguientes: Sexo (H para hombre y M para mujer) y Edad, considerando los siguientes intervalos: [15, 24]; [25, 34]; [35, 44]; [45, 54]; [55, 64]; [65, 74]; [75, 90]

Como uno de los aspectos que nos interesaban era la posible relación entre la observación de algunas características y la edad de las personas, hemos dividido la franja de edad en dos grupos, hasta los 54 (incluidos) y después de los 54. Para esto hemos creado una nueva variable a partir de la variable edad. Esta variable toma el valor 1 si la persona tiene 54 años o menos y el valor 2 si se superan los 54.

Estudios que hemos realizado:

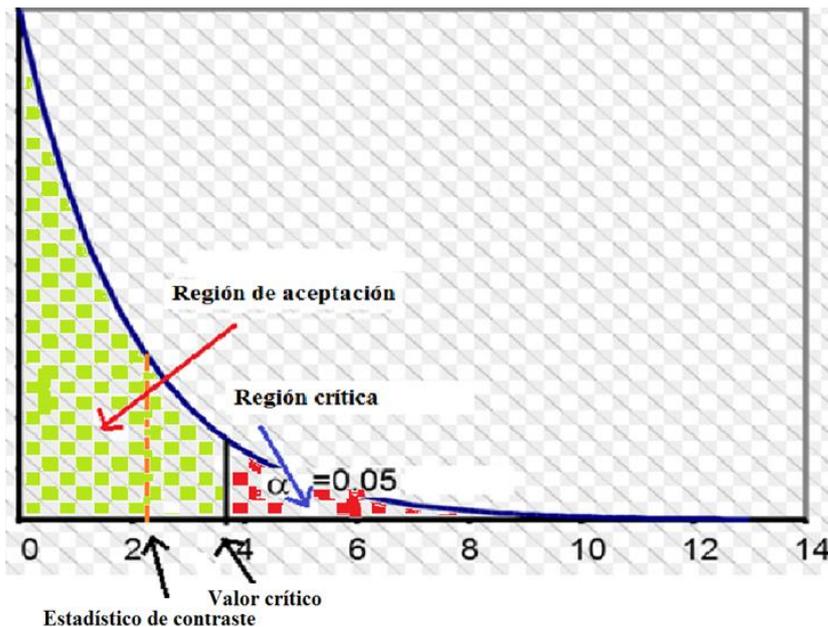
Primero hemos obtenido tablas de frecuencias y representaciones gráficas (diagramas de sectores) para todas las variables con el fin de hacernos una idea de la proporción de síes y noes para cada una de las variables en estudio.

A continuación, hemos intentado establecer una relación entre el hecho de ser zurdo y las variables “cruza brazo derecho sobre izquierdo”; “Cruza pierna izquierda sobre derecha”; “cruza índice izquierdo sobre derecho”. Hemos utilizado tablas de contingencia y el test de la Chi-cuadrado. En cuanto a este test, hemos de decir que, aunque R nos da los resultados y sólo tenemos que fijarnos en el p-valor para llegar a la conclusión, nuestro profesor nos ha indicado que lo hagamos obteniendo las frecuencias esperadas (de manera proporcional, usando reglas de tres) y después, utilizando dichas frecuencias junto a las observadas para obtener el estadístico de contraste (usando una fórmula) y comparando con el valor crítico que viene en las tablas de la chi-cuadrado. En todos los casos hemos utilizado un nivel de significación de 0.05. Si el estadístico de contraste era mayor que el valor crítico o el p-valor que nos daba R era menor que 0.05 rechazábamos la hipótesis nula y concluíamos que había relación entre las variables. La explicación la damos a continuación:

Para obtener el estadístico de contraste usamos la siguiente fórmula:

$$\chi^2 = \sum_{i=1}^n \frac{(f_e - f_o)^2}{f_e}$$

$f_e = \text{frecuencia esperada}$
 $f_o = \text{frecuencia observada}$



Para tomar una decisión se divide la región que queda dibujada debajo de la llamada función chi-cuadrado (curva azul) en dos zonas. Una zona es la llamada región crítica y la otra es llamada región de aceptación. La llamada región crítica tiene un área que vale 0.05. Al valor $\alpha=0.05$ se le llama nivel de significación. Si el estadístico de contraste cae dentro de la región de aceptación no rechazaremos la hipótesis nula. Si cae en la región crítica, rechazaremos la hipótesis nula a favor de la alternativa. La línea que hace de frontera entre las dos regiones es un valor, que depende de α y del tamaño de la muestra, que se busca en unas tablas. Estar dentro de la región de aceptación supone que el estadístico de contraste es menor que el valor crítico que marca la separación entre ambas regiones.

Podemos comparar, también, usando el p-valor, que es el área que está a la derecha de la línea de puntos marrón (ordenada correspondiente al estadístico de contraste). Para no rechazar la hipótesis nula, esta área debe ser mayor que el área correspondiente a la región crítica (recordemos que a esta área se le llama nivel de significación y suele tomar el valor 0.05. Se representa con la letra griega α). Según lo que se observa en el gráfico superior, no rechazaríamos la hipótesis nula.

Otro de nuestros estudios ha ido dirigido a analizar si hay relación entre el hecho de tener los ojos claros (azules o verdes y sus variantes) u oscuros (marrones o negros) y el color del pelo (claro u oscuro). Hemos utilizado el mismo tipo de test que antes.

Considerando que el porcentaje de personas con RH negativo es en España de un 15%, hemos querido comparar si esa proporción se mantiene entre los habitantes de Albacete, utilizando el test para el contraste de la proporción.

Por otro lado, hemos chequeado si hay diferencias entre hombres y mujeres en lo que se refiere a la aparición de las características en estudio. Hemos usado un test para la diferencia de proporciones.

Por último, hemos pensado que quizás podría haber habido una modificación en las características de las personas a lo largo del tiempo. Las sociedades son cada vez más abiertas y tendemos a mezclarnos más con gentes de otras provincias e incluso

países. Esta situación podría haber ocasionado algunos cambios en las características en estudio, por ello, hemos separado la muestra en mayores de 54 años y en menores o iguales de 54 años y hemos comparado proporciones usando, de nuevo, el test para la diferencia de proporciones.

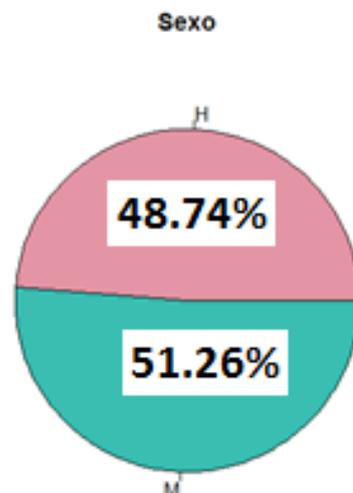
Aclaración importante: Puesto que R ordena las variables cualitativas y sus modalidades alfabéticamente, cuando hacemos comparaciones hay que tener en cuenta que el primer valor de cada variable es el NO y el segundo el SI. Es decir, se está considerando en primer lugar la ausencia del fenotipo. De esto no nos dimos cuenta y empezamos a interpretar mal los resultados y fue el profesor el que nos lo aclaró. Para evitar este problema, hemos realizado los tests para la proporción desde la consola con las instrucciones que nos dio el profesor.

4. Estudios realizados y resultados

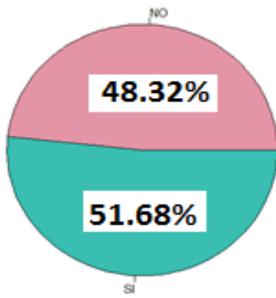
a) Obtención de tablas de frecuencias y gráficos de sectores de todas las variables en estudio.

Para cada una de las variables obtuvimos tablas de frecuencias y porcentajes, así como diagramas de sectores.

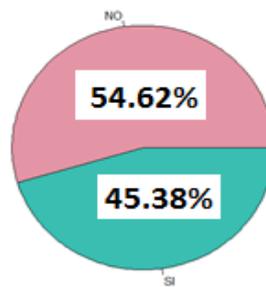
Sexo	Lóbulo de la oreja pegado	Pelos en la 2ª falange	Cruza brazo derecho sobre izquierdo	Es zurdo	Cruza pierna izquierda sobre derecha
H: 116	NO: 163	NO: 150	NO: 115	NO: 189	NO: 147
M: 122	SÍ: 75	SÍ: 88	SÍ: 123	SÍ: 49	SÍ: 91
Cruza dedo índice izquierdo sobre derecho	Dedo índice de menor longitud que anular	Lengua en rollo	Color de pelo claro	Color de ojos claro	RH negativo
NO: 130	NO: 129	NO: 117	NO: 156	NO: 158	NO: 172
SÍ: 108	SÍ: 109	SÍ: 121	SÍ: 82	SÍ: 80	SÍ: 66



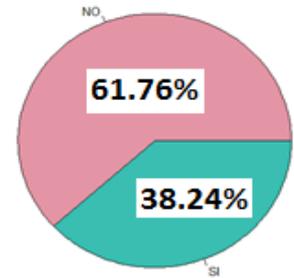
Cruza.brazo.derecho.sobre.izquierdo



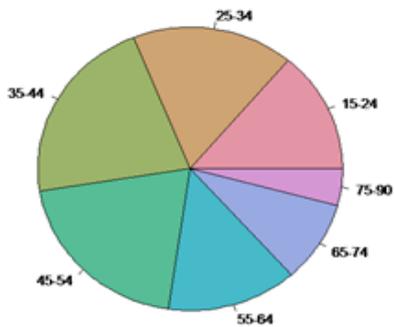
Cruza.indice.Izqui.sobre.derecho



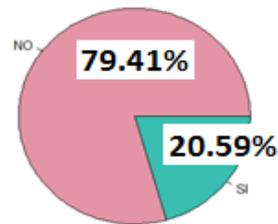
Cruza.Pierna.Izquierda.sobre.derecha



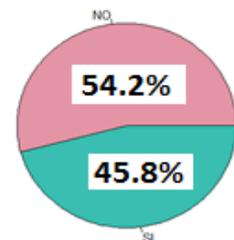
Edad



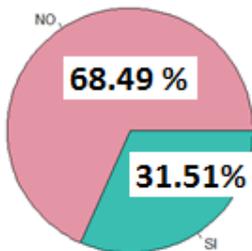
Es.zurdo



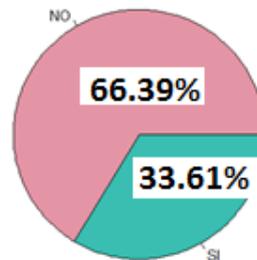
Indice...anular



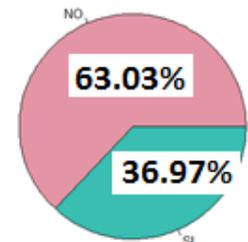
Lóbulo.de.la.oreja.pegado



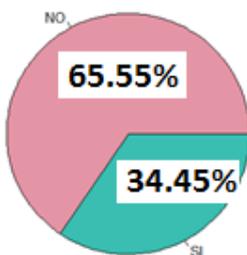
Ojos.claros



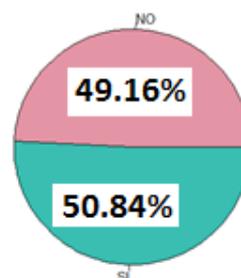
Pelos.en.la.2ª.falange



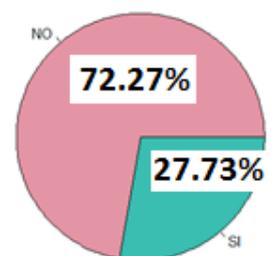
Pelo.claro



X.Lengua.en.rollo



RH.negativo



- b) ¿Hay relación entre las variables “cruza brazo derecho sobre izquierdo”; “Cruza pierna izquierda sobre derecha”; “cruza dedo índice izquierdo sobre derecho” con el hecho de ser zurdo?

Para responder esta pregunta hemos usado tablas de contingencia y el test de la Chi-cuadrado

Como ejemplo de cómo hemos hecho el estudio, vamos a considerar la posible relación entre la variable “cruza brazo derecho sobre izquierdo” y la variable ser zurdo.

	No cruza brazo derecho sobre izquierdo	Sí cruza brazo derecho sobre izquierdo	Total
No es zurdo	96 (189*115/238)=91.32	93 (189*123/238)=97.68	189
Sí es zurdo	19 (49*115/238)=23.68	30 (49*123/238)=25.32	49
	115	123	238

Frecuencias observadas en rojo

Frecuencias esperadas en azul

Estadístico de contraste =2.25

Valor crítico $\chi_{0.05;1x1}^2 = 3.841$

Como el estadístico de contraste es menor que el valor crítico 3.841, consideramos que no hay evidencia en contra de la hipótesis nula, con lo cual, concluimos que no hay relación entre ambas variables. Lo hemos hecho también con R y nos ha dado un p-valor de 0.136 que al ser menor que 0.05, nos lleva a la misma conclusión.

```
Frequency table:
  Cruza.brazo.derecho.sobre.izquierdo
Es.zurdo NO SI
  NO 96 93
  SI 19 30

  Pearson's Chi-squared test

data: .Table
X-squared = 2.2506, df = 1, p-value = 0.1336
```

Para las dos siguientes variables, sólo presentamos aquí los resultados que nos da R.

Relación entre las variables “Cruza pierna izquierda sobre derecha” y “Es zurdo”

```

Frequency table:
      Cruza.Pierna.izquierda.sobre.derecha
Es.zurdo  NO  SI
          NO 129 60
          SI  18 31

      Pearson's Chi-squared test

data: .Table
X-squared = 16.369, df = 1, p-value = 5.212e-05

```

En este caso, rechazaremos la hipótesis nula a favor de la alternativa, con lo cual concluiríamos que ambas variables no son independientes.

Relación entre las variables “Cruza dedo índice izquierdo sobre derecho” y “ser zurdo”

```

Frequency table:
      Cruza.índice.Izqui.sobre.derecho
Es.zurdo  NO  SI
          NO 103 86
          SI  27 22

      Pearson's Chi-squared test

data: .Table
X-squared = 0.0057402, df = 1, p-value = 0.9396

```

Los resultados nos indican que no debemos rechazar la hipótesis nula y que, por lo tanto, ambas variables son independientes.

- c) ¿Hay relación entre el color del pelo y el de los ojos?
Hemos utilizado el mismo procedimiento que en el apartado anterior. Los resultados del test de la chi-cuadrado son los siguientes:

```

Frequency table:
      Ojos.claros
Pelo.claro  NO  SI
          NO 116 40
          SI  42 40

      Pearson's Chi-squared test

data: .Table
X-squared = 12.897, df = 1, p-value = 0.0003292

```

Los resultados nos dicen que sí que hay relación entre las dos variables, ya que el p-valor=0.0003292 es inferior al nivel de significación $\alpha=0.05$

- d) ¿Está la población de Albacete con RH negativo por encima o por debajo de la media nacional que es del 15% para los negativos o no hay diferencias?
Para poder responder usamos el test de la proporción: Primero vamos a ver si es diferente o no. Nuestras hipótesis, son:

$H_0: p=0.15$

$H_1: p \neq 0.15$

$\alpha=0.05$

Los resultados son:
`prop.test(66,238,p=0.15)`

```

1-sample proportions test with continuity correction

data: 66 out of 238, null probability 0.15
X-squared = 29.265, df = 1, p-value = 6.313e-08
alternative hypothesis: true p is not equal to 0.15
95 percent confidence interval:
 0.2223789 0.3395762
sample estimates:
      p
0.2773109

```

Este primer resultado nos indica que la proporción de albaceteños con RH negativo es distinta de la nacional, pero ¿es mayor o menor? Sospechamos que es mayor y lo vamos a chequear con un test unilateral. Las hipótesis son:

$H_0: p \leq 0.15$

$H_1: p > 0.15$

$\alpha=0.05$

Los resultados son:

`prop.test(66,238,p=0.15, alternative="greater")`

```

1-sample proportions test with continuity correction

data: 66 out of 238, null probability 0.15
X-squared = 29.265, df = 1, p-value = 3.157e-08
alternative hypothesis: true p is greater than 0.15
95 percent confidence interval:
 0.2303191 1.0000000
sample estimates:
      p
0.2773109

```

Estos resultados nos indican que hay que rechazar la hipótesis nula a favor de la alternativa, con lo cual, podemos concluir que en la población de Albacete la proporción de personas con RH- es superior a la media nacional.

e) ¿Hay relación entre alguna de las variables y el sexo de las personas?
 Para responder esta pregunta hemos analizado la diferencia entre la proporción de hombres y mujeres para cada uno de los fenotipos:

$H_0: p_H = p_M$

$H_1: p_H \neq p_M$

- Cruza brazo derecho sobre izquierdo. Resultados:

```

Frequency table:
  Cruza.brazo.derecho.sobre.izquierdo
Sexo NO SI
  H 53 63
  M 62 60

```

```
> prop.test(c(63,60),c(116,122))

      2-sample test for equality of proportions with continuity correction

data:  c(63, 60) out of c(116, 122)
X-squared = 0.43806, df = 1, p-value = 0.5081
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.08394389  0.18654422
sample estimates:
   prop 1   prop 2 
0.5431034 0.4918033
```

No hay diferencias entre hombres y mujeres

- Cruza dedo índice izquierdo sobre derecho. Resultados:

```
Frequency table:
Cruza.índice.Izqui.sobre.derecho
Sexo NO SI
H 59 57
M 71 51
```

```
> prop.test(c(57,51),c(116,122))

      2-sample test for equality of proportions with continuity correction

data:  c(57, 51) out of c(116, 122)
X-squared = 1.0116, df = 1, p-value = 0.3145
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.06130336  0.20799641
sample estimates:
   prop 1   prop 2 
0.4913793 0.4180328
```

No hay diferencias entre hombres y mujeres

- Cruza pierna izquierda sobre derecha

```
Frequency table:
Cruza.Pierna.izquierda.sobre.derecha
Sexo NO SI
H 74 42
M 73 49
```

```
> prop.test(c(42,49),c(116,122))

      2-sample test for equality of proportions with continuity correction

data:  c(42, 49) out of c(116, 122)
X-squared = 0.2445, df = 1, p-value = 0.621
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.17133297  0.09219222
sample estimates:
   prop 1   prop 2 
0.3620690 0.4016393
```

No se observan diferencias

- Es zurdo. Resultados

```
Frequency table:
  Es.zurdo
Sexo NO SI
  H  87 29
  M 102 20
```

```
> prop.test(c(29,20),c(116,122))

      2-sample test for equality of proportions with continuity correction

data:  c(29, 20) out of c(116, 122)
X-squared = 2.1933, df = 1, p-value = 0.1386
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.02493412  0.19706527
sample estimates:
 prop 1    prop 2 
0.2500000 0.1639344
```

No hay diferencias

- Dedo índice menor que anular

```
Frequency table:
  Indice...anular
Sexo NO SI
  H  63 53
  M  66 56
```

```
> prop.test(c(53,56),c(116,122))

      2-sample test for equality of proportions with continuity correction

data:  c(53, 56) out of c(116, 122)
X-squared = 2.3737e-30, df = 1, p-value = 1
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.1308747  0.1266350
sample estimates:
 prop 1    prop 2 
0.4568966 0.4590164
```

No se observan diferencias

- Lóbulo de la oreja pegado. Resultados

```
Frequency table:
  Lóbulo.de.la.oreja.pegado
Sexo NO SI
  H  78 38
  M  85 37
```

```
> prop.test(c(38,37),c(116,122))

      2-sample test for equality of proportions with continuity correction

data:  c(38, 37) out of c(116, 122)
X-squared = 0.069643, df = 1, p-value = 0.7919
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.1022024  0.1508174
sample estimates:
  prop 1    prop 2 
0.3275862 0.3032787
```

No se observan diferencias

- Ojos claros. Resultados

```
Frequency table:
  Ojos.claros
Sexo NO SI
  H 76 40
  M 82 40
```

```
> prop.test(c(40,40),c(116,122))

      2-sample test for equality of proportions with continuity correction

data:  c(40, 40) out of c(116, 122)
X-squared = 0.01948, df = 1, p-value = 0.889
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.1115354  0.1454529
sample estimates:
  prop 1    prop 2 
0.3448276 0.3278689
```

No se observan diferencias

- Pelo claro. Resultados

```
Frequency table:
  Pelo.claro
Sexo NO SI
  H 78 38
  M 78 44
```

```
> prop.test(c(38,44),c(116,122))

      2-sample test for equality of proportions with continuity correction

data:  c(38, 44) out of c(116, 122)
X-squared = 0.16013, df = 1, p-value = 0.689
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.16212250  0.09598344
sample estimates:
  prop 1    prop 2 
0.3275862 0.3606557
```

No se observan diferencias

- Pelos en la segunda falange. Resultados

```

Frequency table:
  Pelos.en.la.2ª.falange
Sexo NO SI
  H 57 59
  M 93 29

> prop.test(c(59,29),c(116,122))

      2-sample test for equality of proportions with continuity correction

data:  c(59, 29) out of c(116, 122)
X-squared = 17.583, df = 1, p-value = 2.75e-05
alternative hypothesis: two.sided
95 percent confidence interval:
 0.1442612 0.3975704
sample estimates:
 prop 1    prop 2 
0.5086207 0.2377049

```

Sí se observan diferencias entre hombres y mujeres

El intervalo de confianza nos hace sospechar que la proporción de hombres que tienen pelos en la segunda falange es mayor que la de mujeres. Para chequear nuestras sospechas, vamos a aplicar de nuevo un test para la diferencia de proporciones, pero esta vez unilateral. Hemos de fijarnos en que estamos hablando de diferencias entre hombres y mujeres (en ese orden) y que estamos comparando la proporción de personas que tienen pelos en la segunda falange. Las hipótesis serían:

$$H_0: p_H - p_M \leq 0$$

$$H_a: p_H - p_M > 0$$

```

> prop.test(c(59,29),c(116,122),alternative="greater")

      2-sample test for equality of proportions with continuity correction

data:  c(59, 29) out of c(116, 122)
X-squared = 17.583, df = 1, p-value = 1.375e-05
alternative hypothesis: greater
95 percent confidence interval:
 0.163272 1.000000
sample estimates:
 prop 1    prop 2 
0.5086207 0.2377049

```

Los resultados nos llevan a rechazar la hipótesis nula a favor de la alternativa, con lo cual, se confirma nuestra hipótesis de que la proporción de hombres con pelos en la segunda falange es mayor que la de mujeres.

- RH negativo. Resultados

```

Frequency table:
  RH.
Sexo NO SI
  H 86 30
  M 86 36

> prop.test(c(30,36),c(116,122))

      2-sample test for equality of proportions with continuity correction

data:  c(30, 36) out of c(116, 122)
X-squared = 0.23349, df = 1, p-value = 0.6289
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.15844453  0.08552198
sample estimates:
 prop 1    prop 2 
0.2586207 0.2950820

```

No se observan diferencias

- Lengua en rollo. Resultados

```

Frequency table:
  X.Lengua.en.rollo
Sexo NO SI
  H 62 54
  M 55 67

> prop.test(c(54,67),c(116,122))

      2-sample test for equality of proportions with continuity correction

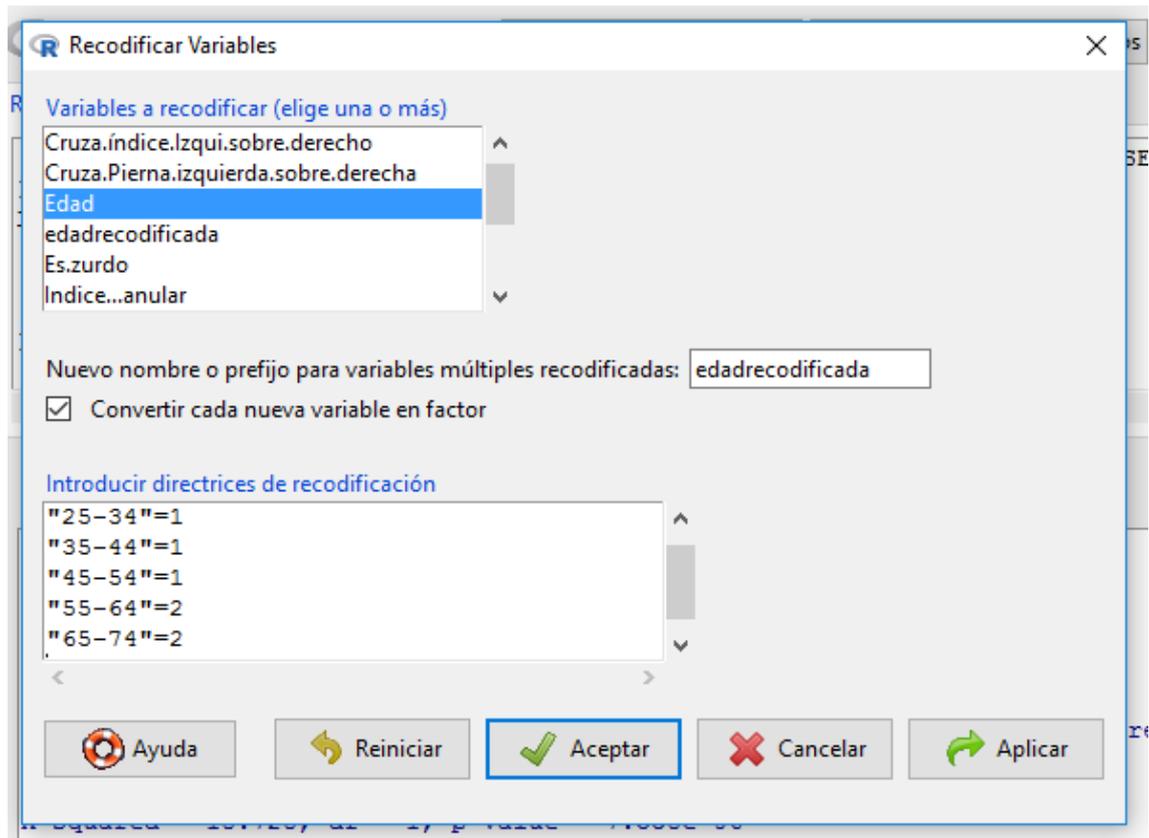
data:  c(54, 67) out of c(116, 122)
X-squared = 1.3474, df = 1, p-value = 0.2457
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.21870241  0.05137623
sample estimates:
 prop 1    prop 2 
0.4655172 0.5491803

```

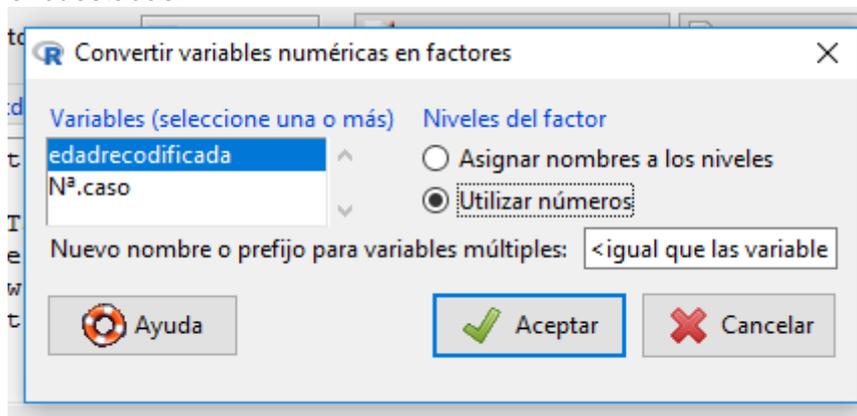
No se observan diferencias

- f) ¿Hay relación entre la observación de algunos fenotipos y la edad de las personas?

Para responder hemos separado las franjas de edad en dos bloques, menores de 54 años y mayores. Para ello hemos de crear una variable que recoja esa información recodificando la variable edad de la siguiente forma:



Después de crear la variable, dado que hemos asignado números, hemos tenido que convertir la variable numérica en factor porque realmente no era una variable numérica, sino algo que nos iba a permitir clasificar a las personas encuestadas.



- Variable cruza brazo derecho sobre izquierdo. Resultados

```
Frequency table:
              Cruza.brazo.derecho.sobre.izquierdo
edadrecodificada NO SI
                1 79 94
                2 36 29
```

```
> prop.test(c(94,29),c(173,65))

      2-sample test for equality of proportions with continuity correction

data:  c(94, 29) out of c(173, 65)
X-squared = 1.4195, df = 1, p-value = 0.2335
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.05520402  0.24960153
sample estimates:
  prop 1    prop 2 
0.5433526 0.4461538
```

No se observan diferencias

- Cruza dedo índice izquierdo sobre derecho. Resultados

```
Frequency table:
              Cruza.índice.Izqui.sobre.derecho
edadrecodificada NO SI
1      92  81
2      38  27
```

```
> prop.test(c(81,27),c(173,65))

      2-sample test for equality of proportions with continuity correction

data:  c(81, 27) out of c(173, 65)
X-squared = 0.34012, df = 1, p-value = 0.5598
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.09875724  0.20440419
sample estimates:
  prop 1    prop 2 
0.4682081 0.4153846
```

No se observan diferencias

- Cruza pierna izquierda sobre derecha. Resultados

```
Frequency table:
              Cruza.Pierna.izquierda.sobre.derecha
edadrecodificada NO SI
1     106  67
2      41  24
```

```
> prop.test(c(67,24),c(173,65))

      2-sample test for equality of proportions with continuity correction

data:  c(67, 24) out of c(173, 65)
X-squared = 0.011164, df = 1, p-value = 0.9159
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.1304914  0.1665963
sample estimates:
  prop 1    prop 2 
0.3872832 0.3692308
```

No se observan diferencias

- Dedo índice menor que anular

```

Frequency table:
                Indice...anular
edadrecodificada NO SI
                1 99 74
                2 30 35

> prop.test(c(74,35),c(173,65))

                2-sample test for equality of proportions with continuity correction

data:  c(74, 35) out of c(173, 65)
X-squared = 1.9084, df = 1, p-value = 0.1671
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.26315297  0.04172122
sample estimates:
 prop 1    prop 2
0.4277457 0.5384615

```

No se observan diferencias

- Ojos claros. Resultados

```

Frequency table:
                Ojos.claros
edadrecodificada NO SI
                1 118 55
                2  40 25

> prop.test(c(55,25),c(173,65))

                2-sample test for equality of proportions with continuity correction

data:  c(55, 25) out of c(173, 65)
X-squared = 0.6667, df = 1, p-value = 0.4142
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.21440313  0.08101051
sample estimates:
 prop 1    prop 2
0.3179191 0.3846154

```

No se observan diferencias

- Pelo claro. Resultados

```

Frequency table:
                Pelo.claro
edadrecodificada NO SI
                1 101 72
                2  55 10

```

```
> prop.test(c(72,10),c(173,65))

      2-sample test for equality of proportions with continuity correction

data:  c(72, 10) out of c(173, 65)
X-squared = 13.26, df = 1, p-value = 0.0002711
alternative hypothesis: two.sided
95 percent confidence interval:
 0.1373507 0.3873270
sample estimates:
 prop 1    prop 2 
0.4161850 0.1538462
```

Se observan diferencias. Apliquemos un test unilateral para ver si hay más personas jóvenes (número 1) con el pelo claro que mayores (número 2).

$H_0: p_1 \leq p_2$

$H_1: p_1 > p_2$

```
> prop.test(c(72,10),c(173,65),alternative="greater")

      2-sample test for equality of proportions with continuity correction

data:  c(72, 10) out of c(173, 65)
X-squared = 13.26, df = 1, p-value = 0.0001355
alternative hypothesis: greater
95 percent confidence interval:
 0.1557441 1.0000000
sample estimates:
 prop 1    prop 2 
0.4161850 0.1538462
```

Los resultados del test nos dicen que la proporción de gente con pelo claro es mayor en las personas más jóvenes.

- Ser zurdo

```
Frequency table:
              Es.zurdo
edadrecodificada NO  SI
1      130      43
2       59       6
```

```
> prop.test(c(43,6),c(173,65))

      2-sample test for equality of proportions with continuity correction

data:  c(43, 6) out of c(173, 65)
X-squared = 6.1318, df = 1, p-value = 0.01328
alternative hypothesis: two.sided
95 percent confidence interval:
 0.05027555 0.26221889
sample estimates:
 prop 1    prop 2 
0.24855491 0.09230769
```

Sí se observan diferencias, por lo tanto, vamos aplica un test unilateral. Sospechamos que hay más zurdos entre las personas jóvenes.

$$p_1 \leq p_2$$

$$p_1 > p_2$$

```
> prop.test(c(43,6),c(173,65),alternative="greater")

      2-sample test for equality of proportions with continuity correction

data:  c(43, 6) out of c(173, 65)
X-squared = 6.1318, df = 1, p-value = 0.006639
alternative hypothesis: greater
95 percent confidence interval:
 0.06561161 1.00000000
sample estimates:
 prop 1    prop 2 
0.24855491 0.09230769
```

El test nos indica que la proporción de zurdos es mayor en las personas jóvenes.

- Pelos en la segunda falange

```
Frequency table:
              Pelos.en.la.2ª.falange
edadrecodificada NO SI
                1 96 77
                2 54 11
```

```
> prop.test(c(77,11),c(173,65))

      2-sample test for equality of proportions with continuity correction

data:  c(77, 11) out of c(173, 65)
X-squared = 14.268, df = 1, p-value = 0.0001586
alternative hypothesis: two.sided
95 percent confidence interval:
 0.1478292 0.4038827
sample estimates:
 prop 1    prop 2 
0.4450867 0.1692308
```

Sí se observan diferencias en lo que se refiere a pelos en la segunda falange.

A continuación, aplicamos un test unilateral sospechando que las personas jóvenes tienen más pelos en la segunda falange que los mayores.

$$p_1 \leq p_2$$

$$p_1 > p_2$$

```
> prop.test(c(77,11),c(173,65),alternative="greater")

      2-sample test for equality of proportions with continuity correction

data:  c(77, 11) out of c(173, 65)
X-squared = 14.268, df = 1, p-value = 7.929e-05
alternative hypothesis: greater
95 percent confidence interval:
 0.1667111 1.0000000
sample estimates:
 prop 1    prop 2 
0.4450867 0.1692308
```

Los resultados nos indican que las personas mayores tienen menos pelos en la segunda falange que las más jóvenes, lo cual está de acuerdo con el hecho de que la piel se va atrofiando según nos hacemos mayores.

- Lóbulo de la oreja pegado. Resultados

```
Frequency table:
              Lóbulo.de.la.oreja.pegado
edadrecodificada NO  SI
1      108    65
2       55    10
```

```
> prop.test(c(65,10),c(173,65))

      2-sample test for equality of proportions with continuity correction

data:  c(65, 10) out of c(173, 65)
X-squared = 9.7738, df = 1, p-value = 0.00177
alternative hypothesis: two.sided
95 percent confidence interval:
 0.09770827 0.34604451
sample estimates:
 prop 1    prop 2 
0.3757225 0.1538462
```

Sí se observan diferencias. Procedamos a aplicar un test unilateral

$$p_1 \leq p_2$$

$$p_1 > p_2$$

```
> prop.test(c(65,10),c(173,65),alternative="greater")

      2-sample test for equality of proportions with continuity correction

data:  c(65, 10) out of c(173, 65)
X-squared = 9.7738, df = 1, p-value = 0.0008851
alternative hypothesis: greater
95 percent confidence interval:
 0.1159698 1.0000000
sample estimates:
 prop 1    prop 2 
0.3757225 0.1538462
```

Podemos decir que la proporción de personas jóvenes con el lóbulo de la oreja pegado a la cara es mayor.

- Lengua en rollo. Resultados

```
Frequency table:
              X.Lengua.en.rollo
edadrecodificada NO  SI
                1  64 109
                2  53  12
```

```
> prop.test(c(109,12),c(173,65))

                2-sample test for equality of proportions with continuity correction

data:  c(109, 12) out of c(173, 65)
X-squared = 35.749, df = 1, p-value = 2.244e-09
alternative hypothesis: two.sided
95 percent confidence interval:
 0.3162344 0.5746505
sample estimates:
 prop 1    prop 2
0.6300578 0.1846154
```

Sí se observan diferencias en la misma línea que en los casos anteriores. Procedamos a aplicar un test unilateral

$$p_1 \leq p_2$$

$$p_1 > p_2$$

```
> prop.test(c(109,12),c(173,65),alternative="greater")

                2-sample test for equality of proportions with continuity correction

data:  c(109, 12) out of c(173, 65)
X-squared = 35.749, df = 1, p-value = 1.122e-09
alternative hypothesis: greater
95 percent confidence interval:
 0.3353062 1.0000000
sample estimates:
 prop 1    prop 2
0.6300578 0.1846154
```

Hay evidencias en contra de la hipótesis nula a favor de la alternativa y podemos concluir que la proporción de personas que pueden sacar la lengua en rollo es mayor entre las personas más jóvenes.

- RH negativo. Resultados

```
Frequency table:
              RH.
edadrecodificada NO  SI
                1 119  54
                2  53  12
```

```

> prop.test(c(54,12),c(173,65))

      2-sample test for equality of proportions with continuity correction

data:  c(54, 12) out of c(173, 65)
X-squared = 3.224, df = 1, p-value = 0.07257
alternative hypothesis: two.sided
95 percent confidence interval:
 4.790027e-05 2.549988e-01
sample estimates:
 prop 1    prop 2 
0.3121387 0.1846154

```

Podemos concluir que no hay diferencias.

5. Conclusiones

- a) Tablas de frecuencias y gráficos de sectores de todas las variables en estudio.

Como nos hemos basado en el censo para obtener el tamaño de la muestra, las proporciones se mantienen, de manera que el número de mujeres es un poco superior al de hombres. Hay mayor cantidad de personas en el grupo de 35 a 44 años seguido del grupo de 45 a 54 años.

Hay más del doble de personas con el lóbulo de la oreja separado y casi la tercera parte de la gente tiene pelos en la segunda falange. El número de personas está más equilibrado en lo que se refiere a cruzar el brazo derecho sobre el izquierdo, así como en lo que se refiere a cruzar el índice izquierdo sobre el derecho. El número de personas que cruza la pierna izquierda sobre la derecha es la mitad del número de personas que no lo hacen. Hay casi cuatro veces más personas diestras que zurdas. Con el dedo índice menor que el anular hay casi la misma gente que con el anular mayor que el índice. En cuanto a poder sacar la lengua doblándola en forma de rollo hay casi la misma proporción de personas que pueden como que no (sin embargo, poder hacerlo es carácter dominante).

La proporción de personas con los ojos oscuros es el doble que con los ojos claros y ocurre lo mismo con el pelo, hay doble de gente con el pelo oscuro que con el pelo claro. Por último, en cuanto al factor RH, el número de personas con RH positivo es caso el triple del número de personas con RH negativo

- b) Relación entre las variables “cruza brazo derecho sobre izquierdo”; “Cruza pierna izquierda sobre derecha”; “cruza dedo índice izquierdo sobre derecho” con el hecho de ser zurdo.

Los resultados nos han llevado a concluir que las variables “cruza brazo derecho sobre izquierdo” y “ser zurdo” son independientes, lo mismo que las variables “cruza dedo índice izquierdo sobre derecho” y “ser zurdo”. Sin embargo, entre las variables “Cruza pierna izquierda sobre derecha” y “ser zurdo” hay relación. Aunque no es, ni mucho menos, prueba de nada, la observación de la tabla de contingencia nos hace pensar que los diestros cruzan menos la pierna izquierda sobre la derecha que los zurdos.

Relación de **independencia** entre variables:

	Cruza brazo derecho sobre izquierdo	Cruza pierna izquierda sobre derecha	Cruza índice izquierdo sobre derecho
Ser zurdo	SÍ	NO	SÍ

c) ¿Hay relación entre el color del pelo y el de los ojos?

Los resultados nos llevan a la conclusión de que ambas variables son dependientes, con lo cual sí que hay relación entre el color del pelo y de los ojos. La tabla de contingencia nos permite ver que hay más gente con pelo y ojos oscuros que gente con pelo oscuro y ojos claros

d) ¿Está la población de Albacete con RH negativo por encima o por debajo de la media nacional que es del 15% para los negativos y del 85% para los positivos?

La respuesta a esta pregunta es que hay evidencias de que la proporción de gente albaceteña con RH- es superior a la media nacional.

e) ¿Hay relación entre alguna de las variables y el sexo de las personas?

No se observan diferencias entre hombres y mujeres excepto en la variable “pelos en la segunda falange”. Concretamente, en esta variable, la realización de un segundo test para la diferencia de proporciones nos lleva a la conclusión de que la proporción de hombres con pelos en la segunda falange es mayor que la de mujeres.

f) ¿Hay relación entre la observación de algunos fenotipos y la edad de las personas?

En cuanto a las variables “Cruza brazo derecho sobre izquierdo”, “Cruza índice izquierdo sobre derecho”, “Cruza pierna izquierda sobre derecha”. “Dedo índice menor que anular”, ojos claros y RH negativo no hay diferencias en cuanto a las proporciones en las dos franjas de edad. Sí las hay en cuanto a tener el lóbulo de la oreja pegado, ya que se concluye que la proporción es mayor en los jóvenes. Algo parecido ocurre con poder sacar la lengua en rollo. La proporción también es mayor en las personas pertenecientes a la primera franja de edad. También hay evidencia de que hay ahora mayor proporción de jóvenes con pelo claro de la que había antes.

Concluimos que las personas mayores tienen menos pelos en la segunda falange que las más jóvenes.

En cuanto a la proporción de zurdos, hay más zurdos entre los jóvenes. Podría haber una explicación en el hecho de que en la actualidad a las personas zurdas no se les impide serlo, mientras que antes, a los niños zurdos se les ataba la mano izquierda para evitar que la usaran, sobre todo al escribir.

	Brazo derecho sobre izquierdo	zurdo	Pierna izquierda sobre derecha	Índice izquierdo sobre derecho	Índice menor que el anular	Factor RH -
Jóvenes		Mayor proporción				
Mayores						

	Lóbulo de la oreja pegado	Pelos en la 2ª falange	Lengua en rollo	Pelo claro	Ojos claros
Jóvenes	Mayor proporción	Mayor proporción	Mayor proporción	Mayor proporción ahora	
Mayores					

6. Posibles mejoras y extensiones del estudio

- Según se dice los españoles somos más altos ahora que antes. Se trataría de verificarlo buscando información sobre la altura de los españoles hace 100 años y comparándola con la de los albaceteños en la actualidad. La muestra estaría formada por hombres y mujeres de 25 a 30 años. La razón de esta franja de edad estaría en que a los 25 hombres y mujeres ya han dejado de crecer y su edad es representativa de lo que ocurre en la actualidad.
- Se podría hacer un estudio del peso medio de los niños de Albacete para compararlo con la media nacional y ver si estamos al mismo nivel, superamos la media o estamos por debajo de ella. Dividiríamos la población en intervalos de edad, empezando a recoger datos a niños desde los 3 años hasta los 18.

