Relación 5a de problemas

1. En una muestra de 15 perros diabéticos sin tratamiento médico se midió el nivel X de glucosa sérica (en mg/dl) obteniéndose una media $\bar{x} = 142.35$ y una varianza $s^2 = 1065.49$.

a) A nivel de significación $\alpha = 0.01$, ¿hay suficiente evidencia muestral de que el nivel esperado de glucosa μ en un perro diabético es superior a 130? Especificar las suposiciones necesarias para resolver el problema. ¿Qué puedes decir (razonadamente) acerca del p-valor del contraste?

b) Obtener, a partir de estos datos, un intervalo de confianza de nivel 0.95 para μ. Calcular el tamaño mínimo muestral que habría que utilizar para, con una probabilidad de 0.95, estimar μ con un error máximo de 5 mg/dl.

2. La aterosclerosis coronaria (AC) es una de las principales causas de mortalidad en los países occidentales. Se cree que la oxidación del colesterol de baja densidad (ver, por ejemplo, Steinberg et al. (1989), *New Engl. J. Med.*, 320, 915-924) es un importante mecanismo en el desarrollo de la AC. Hay una interesante polémica científica sobre los supuestos efectos antioxidantes (y, por tanto, beneficiosos contra la AC) de las bebidas alcohólicas consumidas en cantidades moderadas. Algunos autores han puesto en duda estos efectos cardioprotectores, otros los han atribuido al alcohol en sí mismo y, por último, otros (ver, por ejemplo, Gorinstein et al., 2002, *Nutrition Research* 22, 659-666, para una visión más amplia de este tema) atribuyen la mayor parte de los efectos beneficiosos a las sustancias fenólicas que están contenidas en el vino tinto en mucha mayor medida que en otras bebidas alcohólicas. Según los defensores de esta última teoría, el vino tinto (consumido siempre en cantidades muy moderadas) sería mucho más cardiosaludable que las demás bebidas alcohólicas.

Se han determinado los valores de epicatequina (una sustancia fenólica) en 10 muestras de vino tinto, encontrando que la media muestral era 195.1 mg/l y el error típico 10.051. Los correspondientes valores para 10 muestras de cerveza fueron 65.5 mg/L y 3.4184. En vista de estos resultados ¿puede aceptarse que las varianzas de la epicatequina en el vino tinto y en la cerveza son iguales?, ¿puede decirse que son significativamente diferentes los contenidos medios de esta sustancia en el vino tinto y en la cerveza?

Responde a la misma pregunta para el caso del ácido ferúlico en el que los resultados obtenidos fueron, para el vino tinto, 7.2 mg/l (media) y 0.4541 mg/l (error típico) (tamaño muestral = 10) y para la cerveza 6.8 mg/l (media) y 0.3571 mg/l (error típico) (tamaño muestral = 10).

3. Se sospecha que ciertos riesgos ambientales y laborales pueden alterar la proporción de nacimientos de varones. En particular, hay controversia respecto al efecto que la exposición a radiación ionizante tiene sobre el *sex ratio* o cociente del número de nacimientos masculinos respecto a aquél de nacimientos femeninos. Scherb *et al.* (2013)[1] estudiaron el efecto sobre el *sex ratio* del accidente en la Central Nuclear de Chernóbil (Ucrania) en abril de 1986. Concretamente, desde 1959 hasta 1986 (ambos años incluidos) nacieron en Rusia 31 706 752 varones de un total de 61 890 823 nacimientos. Desde 1987 hasta el 2010 (ambos años incluidos) en Rusia nacieron 38 360 531 bebés de los que 19 727 605 fueron varones.

[1]Fuente: Scherb, H. Kusmierz, R. y Voigt, K. (2013). Increased sex ratio in Russia and Cuba after Chernobyl: a radiological hypothesis. *Environmental Health*, 12:63.

a) A nivel $\alpha = 0.05$, ¿hay suficiente evidencia estadística de que el accidente de Chernóbil redujo la proporción de nacimientos de niñas (sobre el total de nacimientos) en Rusia?.

b) Explica el siguiente código de R e interpreta la correspondiente salida:

```
prop.test(c(31706752,19727605),c(61890823,38360531),alternative="less",
           correct=FALSE)

        2-sample test for equality of proportions without continuity
   correction

data:  c(31706752, 19727605) out of c(61890823, 38360531)
X-squared = 366.74, df = 1, p-value < 2.2e-16
alternative hypothesis: less
95 percent confidence interval:
 -1.000000000 -0.001798013
sample estimates:
   prop 1    prop 2
0.5123013 0.5142683
```

c) Construye un intervalo de confianza al 95 % para la proporción de nacimientos de varones sobre el total de nacimientos después del accidente nuclear. A nivel de significación del 5 %, ¿hay evidencia a favor de la hipótesis de que dicha proporción es diferente de la proporción análoga a nivel mundial, que es de 0.517?

4. Se mide el grado X de expresión de un gen en el tejido ovárico de 23 mujeres sanas[2], obteniéndose los siguientes datos:

0.51	0.52	0.62	0.67	0.67	0.70	0.76	0.76	0.79	0.81	0.81	0.84
0.89	0.94	1.01	1.09	1.15	1.15	1.16	1.27	1.35	1.37	2.63	

Para estos datos se obtiene $\sum_{i=1}^{23} x_i = 22.4736$ y $\sum_{i=1}^{23} x_i^2 = 26.2135$.

a) Se mide el grado de expresión del mismo gen en el tejido ovárico de 30 mujeres con cáncer de ovario, obteniéndose la siguiente muestra:

0.81	0.70	0.64	0.67	0.60	0.42	0.70	0.55	0.98	1.10
0.69	0.34	0.60	0.49	1.19	0.87	2.33	1.16	0.50	0.95
0.81	2.78	1.25	0.69	1.03	0.69	0.57	0.72	0.72	0.94

cuya media y desviación típica son 0.88 y 0.51 respectivamente. A nivel $\alpha = 0.05$, ¿hay suficiente evidencia muestral de que el nivel esperado de expresión de ese gen es diferente en mujeres sanas y en pacientes con cáncer de ovario?

b) El p-valor del contraste ¿es mayor o menor que 0.05? Razonar la respuesta.

5. En el paquete `PairedData` de `R` aparece el conjunto de datos `Shoulder`. Se trata de la flexibilidad (medida en grados según el rango de movilidad del brazo) de los hombros izquierdo y derecho en un grupo de 15 nadadores (`Swimmer`) y en otro grupo de 15 personas sedentarias (`Control`). A continuación reproducimos los datos con `R`, así como algunos cálculos realizados sobre los mismos:

[2]Fuente de los datos: Pepe *et al.* (2003). Selecting Differentially Expressed Genes from Microarray Experiments. *Biometrics*, 59,133–142.

```
library("PairedData")
Shoulder

  Subject   Group Left Right
1       S1 Swimmer  193   192
2       S2 Swimmer  208   207
3       S3 Swimmer  198   198
4       S4 Swimmer  201   203
5       S5 Swimmer  196   194
6       S6 Swimmer  196   193
7       S7 Swimmer  211   214
8       S8 Swimmer  206   207
9       S9 Swimmer  197   195
10     S10 Swimmer  204   198
11     S11 Swimmer  197   198
12     S12 Swimmer  205   206
13     S13 Swimmer  207   200
14     S14 Swimmer  204   204
15     S15 Swimmer  204   205
16     S16 Control  184   197
17     S17 Control  172   187
18     S18 Control  178   180
19     S19 Control  186   175
20     S20 Control  194   192
21     S21 Control  188   189
22     S22 Control  164   185
23     S23 Control  202   168
24     S24 Control  182   204
25     S25 Control  186   181
26     S26 Control  188   182
27     S27 Control  186   172
28     S28 Control  192   183
29     S29 Control  204   189
30     S30 Control  178   198
```

```
Nadador = (Shoulder$Group=="Swimmer")

HombrosNadador=Shoulder[Nadador,c("Left","Right")]
HombrosControl=Shoulder[!Nadador,c("Left","Right")]

mean(HombrosNadador$Left)
[1] 201.8
mean(HombrosNadador$Right)
[1] 200.9333
mean(HombrosControl$Left)
[1] 185.6
mean(HombrosControl$Right)
[1] 185.4667
var(HombrosNadador$Left)
[1] 28.45714
var(HombrosNadador$Right)
[1] 39.49524
var(HombrosControl$Left)
[1] 108.1143
var(HombrosControl$Right)
[1] 97.69524
```

a) A nivel de significación $\alpha = 0.05$, ¿hay diferencia entre las flexibilidades esperadas del hombro derecho e izquierdo en un nadador? Especifica las suposiciones previas para la resolución.

b) Determina un intervalo de confianza al 95 % para la flexibilidad esperada en el hombro derecho de un nadador. A nivel de significación $\alpha = 0.05$, ¿hay evidencia estadística a favor de la hipótesis de que dicha flexibilidad media es diferente de 200 grados?. Especifica las suposiciones previas para resolver este apartado.

c) Explica e interpreta el siguiente código y salida del programa R:

```
t.test(HombrosNadador$Right,mu=200,alternative="greater")

        One Sample t-test

data:  HombrosNadador$Right
t = 0.57519, df = 14, p-value = 0.2871
alternative hypothesis: true mean is greater than 200
95 percent confidence interval:
 198.0753      Inf
```

```
sample estimates:
mean of x
 200.9333
```

6. Se realiza un experimento para comparar los incrementos en los niveles plasmáticos de insulina producidos por la ingesta de carne y de pescado. Para ello se midieron los incrementos (medidos en picomoles por litro) producidos en la concentración de insulina en la sangre de 6 voluntarios, 90 minutos después de comer un bistec de 250 gr. Dos días más tarde se realizó de nuevo el experimento con las mismas 6 personas, después de consumir un filete de pescado. En la siguiente tabla se indican los respectivos incrementos en la concentración de insulina producidos por la carne y por el pescado:

Persona:	1	2	3	4	5	6
Resultados con la carne:	109	106	111	105	110	108
Resultados con el pescado:	100	95	105	106	80	88

a) Proporcionan estos datos, suficiente evidencia estadística, al nivel de significación 0.05, para afirmar que el incremento medio en la concentración de insulina producido por el consumo de carne es mayor que el producido por el consumo de pescado? Responde a la misma pregunta anterior utilizando un nivel de significación de 0.01. Indica las suposiciones previas necesarias para garantizar la validez del procedimiento empleado.

b) Calcula el tamaño muestral que sería necesario para estimar el incremento medio producido por el consumo de carne, de manera que se tenga una probabilidad 0.95 de cometer, como máximo, un error de 0.2 unidades.

7. Se ha comprobado que la probabilidad de curación espontánea (sin medicación alguna) de cierta enfermedad es de 0.4. Un laboratorio ha obtenido un nuevo medicamento para tratar la enfermedad y quiere demostrar que cura la enfermedad en más de un 40 % de los casos. Para ello se aplica el tratamiento a 100 pacientes que sufren la enfermedad y se observa cuántos de ellos se curan.

a) Si se han curado 50 personas de las 100, a nivel $\alpha = 0.05$, ¿puede afirmarse que el medicamento tiene una probabilidad de curar la enfermedad por encima de 0.4? Calcula el p-valor del contraste.

b) ¿Cuántas personas de las 100 deberían curarse como mínimo para poder afirmar al nivel $\alpha = 0.001$ que la probabilidad de curación con el nuevo tratamiento supera el 40 %?

c) Supongamos que la probabilidad de curación con el tratamiento fuese realmente de 0.5 y que se realiza el test de nivel 0.05 con 100 personas. ¿Cuál sería la probabilidad de error, es decir, la probabilidad de rechazar el medicamento como inútil?.

8. La gerente de un hospital privado afirma que, durante los fines de semana, el tiempo medio de espera en urgencias es de 10 minutos. Basándose en su experiencia personal, un facultativo afirma, por el contrario, que el tiempo medio que un paciente de urgencias espera, durante el fin de semana, hasta que es atendido es superior a lo que dice la dirección del hospital. El facultativo recoge durante varios fines de semana los tiempos de espera de 40 pacientes elegidos al azar en el servicio de urgencias. El promedio de tiempos de espera de los 40 pacientes es 11 minutos con una desviación típica (s_{n-1}) de 3 minutos. A nivel de significación del 5 %, ¿hay suficiente evidencia muestral a favor de la opinión del facultativo? ¿Qué puedes decir sobre el p-valor del contraste?