

Examen parte Específica curso 2021 - 2022

Juan de Dios Luna del Castillo



**UNIVERSIDAD
DE GRANADA**

Todo el material para el conjunto de actividades de este curso ha sido elaborado y es propiedad intelectual del grupo **BioestadísticaR** formado por:

Antonio Martín Andrés
Juan de Dios Luna del Castillo,
Pedro Femia Marzo,
Miguel Ángel Montero Alonso,
Christian José Acal González,
Pedro María Carmona Sáez,
Juan Manuel Melchor Rodríguez,
José Luis Romero Béjar,
Manuela Expósito Ruíz,
Juan Antonio Villatoro García.

Todos los integrantes del grupo han participado en todas las actividades, en su elección, construcción, correcciones o en su edición final, no obstante, en cada una de ellas, aparecerán uno o más nombres correspondientes a las personas que han tenido la máxima responsabilidad de su elaboración junto al grupo de **BioestadísticaR**.

Todos los materiales están protegidos por la Licencia Creative Commons **CC BY-NC-ND** que permite "descargar las obras y compartirlas con otras personas, siempre que se reconozca su autoría, pero no se pueden cambiar de ninguna manera ni se pueden utilizar comercialmente".

Este cuestionario consta de 40 preguntas de opción múltiple y cuatro problemas con 15 preguntas cada uno, sin orden preestablecido, cada cuestión correcta se puntúa con 0.5 sin penalización si es incorrecta.

La navegación es libre, se puede avanzar y retroceder las veces que sea necesario.

El tiempo de realización es de **2 horas y 15 minutos**, una vez transcurrido se cerrará automáticamente.

Pregunta 1 Enunciado de la pregunta

Se llevó a cabo un estudio para ver si el porcentaje de mortalidad de pacientes ingresados en UCI a los 6 meses del alta era diferente en los pacientes tratados con el suero salino clásico (SS) y en pacientes tratados con suero multielectrolítico (M). Hecho el estudio con dos muestras de pacientes se obtuvieron los siguientes resultados: 530 de 2433 pacientes (21.8%) en el grupo SM y 530 de 2413 pacientes (22.0%) en el grupo SS. A la vista de estos resultados los investigadores concluyeron: "No hay diferencias entre grupos, por lo que no tenemos evidencia de que la probabilidad de morir sea diferente con uno u otro suero". Diga si la afirmación es correcta o no y por qué.

Fuente: DOI: 10.1056/NEJMoa2114464. Solo ha servido como inspiración y sus resultados pueden ser diferentes de las afirmaciones hechas en la pregunta

- a. La afirmación es incorrecta por qué falta el error β y el error α al que se hizo el test.
- b. Ninguna de las otras alternativas es correcta
- c. La afirmación es incorrecta por qué falta el error β al que se hizo el test.
- d. La afirmación es incorrecta por qué falta el error α al que se hizo el test.

Pregunta 2 Enunciado de la pregunta

Señale el modelo de regresión lineal simple para las variables x (independiente) e y (dependiente) que sea correcto.

- a. $x \rightarrow N(\alpha + \beta \cdot y; \beta \cdot y)$ $x \rightarrow N(\alpha + \beta \cdot y; \beta \cdot y)$
- b. $x \rightarrow N(\alpha + \beta \cdot y; \sigma)$ $x \rightarrow N(\alpha + \beta \cdot y; \sigma)$
- c. $y \rightarrow N(\alpha + \beta \cdot x; \sigma)$ $y \rightarrow N(\alpha + \beta \cdot x; \sigma)$
- d. $y_2 \rightarrow N(\alpha + \beta \cdot x; \beta \cdot x)$ $y_2 \rightarrow N(\alpha + \beta \cdot x; \beta \cdot x)$

Pregunta 3 Enunciado de la pregunta

La fuerza de la relación entre dos variables medida por el coeficiente de correlación de Pearson es $r = -0.60$. Elija la opción correcta de entre las que aparecen como alternativas.

- a. El porcentaje de variabilidad de la y que es capaz de explicar la x es de un -16%

- b. El porcentaje de variabilidad de la y que es capaz de explicar la x es de un 36%
- c. El porcentaje de variabilidad de la y que es capaz de explicar la x es de un -36%
- d. El porcentaje de variabilidad de la y que es capaz de explicar la x es de un 16%

Pregunta 4 Enunciado de la pregunta

Elija el método de medida más preciso entre los cuatro que se presentan con los resultados de 10 medidas de cuatro muestras diferentes.

- a. Método C: $\bar{a}=300s=60$ $\bar{a}=300s=60$
- b. Método A: $\bar{a}=100s=10$ $\bar{a}=100s=10$
- c. Método B: $\bar{a}=200s=30$ $\bar{a}=200s=30$
- d. Método D: $\bar{a}=200s=40$ $\bar{a}=200s=40$

Pregunta 5 Enunciado de la pregunta

La media muestral de la concentración de colesterol en sangre es un estimador puntual de la media poblacional de la concentración de colesterol en sangre. Esa media muestral es:

- a. Un conjunto de valores que contiene en su interior el verdadero valor de la media muestral.
- b. Una variable aleatoria que tiene hasta dos valores diferentes dependiendo de la muestra que se tome.
- c. Una variable aleatoria puesto que su valor variaría de muestra a muestra.
- d. Una constante ya que no cambia y su valor es siempre fijo.

Pregunta 6 Enunciado de la pregunta

Se llevó a cabo un estudio para ver si el porcentaje de mortalidad de pacientes ingresados en UCI a los 6 meses del alta era diferente en los pacientes tratados con el suero salino clásico (SS) y en pacientes tratados con suero multielectrolítico (SM). Hecho el estudio con dos muestras de pacientes se obtuvieron los siguientes resultados: 530 de 2433 pacientes murieron (21.8%) en el grupo SM y 530 de 2413 pacientes murieron (22.0%) en el grupo SS y además la diferencia de ambas proporciones valía -0.15% con un intervalo, al 95%, que va desde -3.60% a +3.30% y hecho el contraste de hipótesis se obtuvo $P=0.90$.

Ante estos resultados, ha de tomarse una decisión fiable, si es posible. Elija la opción correcta de entre las cuatro.

Fuente: DOI: 10.1056/NEJMoa2114464. Solo ha servido como inspiración y sus resultados pueden ser diferentes de las afirmaciones hechas en la pregunta.

- a. Como $P=0.90$, muy grande, no rechazamos la hipótesis nula. De los resultados se puede concluir que esa no significación se debe a que las diferencias entre ambas proporciones son moderadas y también al tamaño de muestra (no son muy grandes). Luego si la diferencia de interés es mayor, en valor absoluto, de 3.30 la decisión sería fiable ya que el tamaño de muestra no importa.
- b. Como $P=0.90$, muy grande, no rechazamos la hipótesis nula. De los resultados se puede concluir que esa no significación se debe a que las diferencias entre ambas proporciones son pequeñas y también al tamaño de muestra (no son muy grandes). Luego si la diferencia de interés es mayor, en valor absoluto, de 3.60 la decisión sería fiable ya que el tamaño de muestra no importa.
- c. Como $P=0.90$, muy grande, no rechazamos la hipótesis nula. De los resultados se puede concluir que esa no significación se debe a que las diferencias entre ambas proporciones son pequeñas y no al tamaño de muestra (que son grandes). Luego si la diferencia de interés es menor, en valor absoluto, de 3.60 la decisión sería fiable.
- d. Como $P=0.90$, muy grande, no rechazamos la hipótesis nula. De los resultados se puede concluir que esa no significación se debe a que las diferencias entre ambas proporciones son pequeñas y no al tamaño de muestra (que son grandes). Luego si la diferencia de interés es mayor, en valor absoluto, de 3.60 la decisión sería fiable.

Pregunta 7 Enunciado de la pregunta

Señale en cuál de las 4 muestras que se presentan la media representa mejor a las observaciones de ella.

- a. Muestra 4: $n=100$ $\bar{x}=320.3$ $s^2=34.4$ $\bar{x}=320.3$ $s^2=34.4$
- b. Muestra 3: $n=100$ $\bar{x}=320.3$ $s^2=34.1$ $\bar{x}=320.3$ $s^2=34.1$
- c. Muestra 1: $n=100$ $\bar{x}=320.3$ $s^2=33.1$ $\bar{x}=320.3$ $s^2=33.1$
- d. Muestra 2: $n=100$ $\bar{x}=320.3$ $s^2=33.9$ $\bar{x}=320.3$ $s^2=33.9$

Pregunta 8 Enunciado de la pregunta

Hoy día la mortalidad por COVID19 entre las personas que sufren la infección es un suceso que ocurre en menos de un 10% de los casos.

Para estudiar la asociación entre la mortalidad en pacientes con COVID19 y la obesidad de los mismos se llevó a cabo un estudio, que denominaremos a partir de ahora A, en el que se tomó una muestra de 100 muertos por la enfermedad y se vio que 28 eran obesos

y de otra muestra de 200 pacientes que no murieron de la enfermedad 11 eran obesos. A la vista de estos resultados: a) ¿Existe relación entre la obesidad y la mortalidad por COVID19? b) Estímese, si es posible, el riesgo relativo de la obesidad para la muerte por COVID19.

Solución:

(Se presenta ahora un informe que usted debe rellenar. El rellenado se hará unas veces con una elección múltiple, otras con un número y otras con un texto libre; siga las instrucciones sobre el formato de introducción de la información rigurosamente.)

Apartado a)

Para emitir el informe empezaremos indicando el problema, la enfermedad, que en este caso es:

El potencial factor de riesgo es:

El grupo de riesgo son:

El tipo de estudio que se ha hecho y que nos ayuda a responder al ejercicio es:

Con ese tipo de estudio el test de independencia en este caso tendría la hipótesis nula siguiente:

La instrucción de R, de la librería BioestadísticaR, que hemos de aplicar sería:

Los resultados figuran en el recuadro siguiente:

```
# Análisis de tablas 2x2
# -----

# Frecuencias observadas
      Obeso SI   Obeso NO Total
Muerte SI      31      72   103
Muerte NO      15     189   204
Total           46     261   307

# Test Chi-cuadrado para un estudio retrospectivo

Frecuencia mínima esperada = 15.4332 > 7.7

X2 = 27.7821,  gl = 1,  p < 0.001, (cpc = 1)
Test exacto de Fisher: p < 0.001

--- Otros criterios X2:
X2 = 27.7937,  gl = 1,  p < 0.001, (sin cpc)
X2 = 26.0369,  gl = 1,  p < 0.001, (cpc de Yates = 153.5)

# Medidas de asociación para un estudio retrospectivo

Riesgo atribuible*:
Ra=0.5497; 95%-IC(Ra)= (0.3132, 0.7048)
* La estimación de Ra para estudios retrospectivos es una aproximación válida si la prevalencia de la enfermedad es baja: P(E) < 10%

Razón del producto cruzado (odds ratio):
OR=5.425; 95%-IC(OR)= (2.7301, 10.3353)
* La estimación para OR sirve de aproximación al riesgo relativo siempre que la prevalencia de la enfermedad sea P(E) < 10%
```

Para hacer el test chi-cuadrado ha de verificarse las condiciones de validez, que se verifican en este caso puesto que la frecuencia mínima esperada vale: (cantidad redondeada a un decimal): (> 7.7.

Hecho el test chi-cuadrado (con cpc=1 no se considere ni sin cpc ,ni el de cpc de Yates, ni el test exacto de Fisher) su valor es(máximo un decimal): (con (g.l.

P(Máximo 3 decimales y poner 0.001 siempre que p<0.001): ().

En función del valor de P podemos . Luego la mortalidad por COVID19 y la obesidad .

A partir de aquí la estimación puntual del riesgo relativo será (máximo redondeado con un decimal): () y la estimación por intervalos, al(un valor entero de 0 a 100): (,) dará dos valores que son (máximo redondeado con un decimal): (,).

Pregunta 9 Enunciado de la pregunta

Señale la pareja de valores de la pendiente y del coeficiente de correlación que es compatible de entre las cuatro que se le ofrecen

- a. $b=-0.03$, $r=0$
- b. $b=-0.03$, $r=-0.80$
- c. $b=-0.03$, $r=0.80$
- d. $b=0$, $r=-0.80$

Pregunta 10 Enunciado de la pregunta

En un test de comparación de dos medias, el estadístico de contraste empleado fue la diferencia de ambas medias. La región de aceptación de dicho test, al 5% de error, iba de 1.31 a 12.18. En un estudio concreto con dos muestras el valor del estadístico de contraste fue de 8.31 ¿Qué decisión se tomaría y por qué?

- a. Como el valor del estadístico de contraste cae en la región de aceptación, es un resultado que no es improbable supuesta cierta la hipótesis nula, por ello no podremos rechazar la hipótesis nula.
- b. Como el estadístico cae en la región de aceptación que es un conjunto de valores que ocurre con poca probabilidad supuesta cierta la hipótesis nula no nos queda más remedio que rechazar la hipótesis nula.
- c. El estadístico de contraste cae en la región crítica que es una región improbable supuesta cierta la hipótesis nula, por ello se rechaza la hipótesis nula.
- d. Independientemente de dónde caiga el estadístico de contraste, el valor 8.31 es raro siendo cierta la hipótesis nula por lo que hemos de rechazar dicha hipótesis.

Pregunta 11 Enunciado de la pregunta

Señale la opción correcta de entre las 4 siguientes.

- a. La anchura de un intervalo será tanto menor conforme la confianza sea menor, fijado el tamaño de muestra.
- b. La anchura de un intervalo será tanto menor conforme la confianza sea mayor, fijado el tamaño de muestra.
- c. La anchura de un intervalo será tanto menor conforme la confianza sea mayor, independientemente del tamaño de muestra
- d. La anchura de un intervalo será tanto menor conforme la confianza sea mayor, independientemente del tamaño de la muestra

Pregunta 12 Enunciado de la pregunta

Una persona A que está hospitalizada por COVID19 tiene una probabilidad de morir, durante la estancia, de 0.6 y otra persona, B, tiene una probabilidad de morir, durante la estancia hospitalaria, de 0.1, ¿Cuál debería preocupar más, por el riesgo de morir que tiene?

- a. La A y la B tienen el mismo riesgo de morir ya que en ambos los resultados posibles son morir o no morir, luego en ellas la probabilidad sería $1/2=0.5$
- b. La B ya que tiene menos probabilidad de morir que de no morir
- c. La A ya que, si tuviéramos infinitas personas en su situación, el 60% de ellas morirían en la estancia hospitalaria, mientras que sólo lo harían el 10% de las personas en las mismas circunstancias que la B
- d. La A ya que tiene 3 veces más probabilidad de morir que de no morir.

Pregunta 13 Enunciado de la pregunta

Elija la definición correcta del error α y del error β

- a. $\alpha=P(\text{decidir } H_0 | \text{cierta } H_1)$ y $\beta=P(\text{decidir } H_1 | \text{cierta } H_1)$
- b. $\alpha=P(\text{decidir } H_0 | \text{cierta } H_1)$ y $\beta=P(\text{decidir } H_0 | \text{cierta } H_0)$
- c. $\alpha=P(\text{decidir } H_1 | \text{cierta } H_0)$ y $\beta=P(\text{decidir } H_0 | \text{cierta } H_1)$
- d. $\alpha=P(\text{decidir } H_0 | \text{cierta } H_0)$ y $\beta=P(\text{decidir } H_1 | \text{cierta } H_1)$

Pregunta 14 Enunciado de la pregunta

Ante la frase " Hecho el test se obtuvo $P=0.4434$ ", ¿Cuál sería la conclusión? Elija la opción más correcta

- a. Siendo la probabilidad de obtener un resultado tan extraño o más que el obtenido, supuesta cierta la hipótesis nula, $P=0.4434$ que es de los considerados pequeños podemos rechazar la hipótesis nula.
- b. Siendo la probabilidad de obtener un resultado tan extraño o más que el obtenido, supuesta cierta la hipótesis nula, $P=0.4434$ que es de los considerados no suficientemente pequeños debemos aceptar la hipótesis nula, pero deberíamos aumentar el tamaño de muestra
- c. La decisión no se puede tomar porque con los datos que tenemos no sabemos si será o no fiable

d. Siendo la probabilidad de obtener un resultado tan extraño o más que el obtenido, supuesta cierta la hipótesis nula, $P=0.4434$ que es de los considerados grandes debemos aceptar la hipótesis nula.

Pregunta 15 Enunciado de la pregunta

Con el objeto de ver la efectividad de una vacuna contra el COVID19 se tomó una muestra de personas vacunadas contra el virus y otra de personas no vacunadas; a ambas se las siguió durante 3 meses y se vio las que se infectaban y las que no se infectaban, ¿Qué tipo de estudio observacional es ese?

- a. Ninguno de los anteriores
- b. Prospectivo
- c. Retrospectivo o de casos y controles
- d. Transversal

Pregunta 16 Enunciado de la pregunta

Un estudio para determinar si había diferencias por término medio en el nivel de hemoglobina glicosilada (HbA1c) tras seguir un tratamiento no-farmacológico en pacientes diabéticos Tipo II, se llevó de dos maneras diferentes. Manera 1: En la que a una muestra de 10 pacientes se les midió la HbA1c Antes del tratamiento y Tras el tratamiento, ambas variables están relacionadas de manera positiva. Manera 2: Dos muestras, de 10 individuos cada una, y que partían de valores basales de HbA1c prácticamente iguales, midiéndoseles a unos la HbA1c tras el tratamiento y a otros tras no aplicarles ningún tratamiento. En las dos situaciones se obtuvo que la HbA1c sin tratamiento dio una media de 7.15 y con tratamiento la media fue de 6.25. ¿En cuál de los dos casos el test de comparación de medias dará un valor de P más bajo? Elija la opción correcta de entre las cuatro

- a. En la Manera 2, ya que en ese caso son muestras independientes y el test con muestras independientes, para los mismos tamaños de muestra, es más potente que el test de muestras apareadas.
- b. En la Manera 1, ya que en ese caso son muestras apareadas y el test con muestras apareadas, para los mismos tamaños de muestra, es más potente que el test de muestras independientes.
- c. En la Manera 2, ya que en ese caso son muestras apareadas y el test con muestras apareadas, para los mismos tamaños de muestra, es más potente que el test de muestras independientes.

d. En la Manera 1, ya que en ese caso son muestras independientes y el test con muestras independientes, para los mismos tamaños de muestra, es más potente que el test de muestras apareadas.

Pregunta 17 Enunciado de la pregunta

Elija la razón correcta para hacer el test de homogeneidad de varianzas previo al del test de comparación de medias, en el caso de muestras independientes y variables aleatorias normales, a un nivel de error α superior al clásico del 5%.

a. Por qué, siendo este un test auxiliar, no el definitivo, tan grave es cometer el error de Tipo I como el de Tipo II, de forma que tomando el error α mayor hacemos menor el error β y nos cubrimos así del error de Tipo II.

b. Por qué, siendo este un test auxiliar, no el definitivo, más grave es cometer el error de Tipo II, de forma que tomando el error α menor nos cubrimos de dicho error.

c. Por qué, siendo este un test auxiliar, no el definitivo, más grave es cometer el error de Tipo I, de forma que tomando el error α mayor nos cubrimos de dicho error.

d. Por qué, siendo este un test auxiliar, no el definitivo, tan grave es cometer el error de Tipo I como el de Tipo II, de forma que tomando el error α mayor hacemos igual el error β y nos cubrimos así del error de Tipo II.

Pregunta 18 Enunciado de la pregunta

Se prueba la efectividad contra el COVID19 en una muestra 300 personas hospitalizadas con dicha enfermedad de 30 hospitales de Europa. Se hace con estos pacientes porque son los que están disponibles y estamos seguros de que tienen la enfermedad. ¿A quién se podrían aplicar los resultados de tal prueba? Elija la opción correcta.

a. Sólo a los ingresados porque a los no ingresados no hace falta aplicarles el tratamiento.

b. Sólo a los ingresados que no son una muestra aleatoria de los pacientes con COVID19

c. A todos los pacientes con COVID19 ya que los no ingresados también están representados con ellos

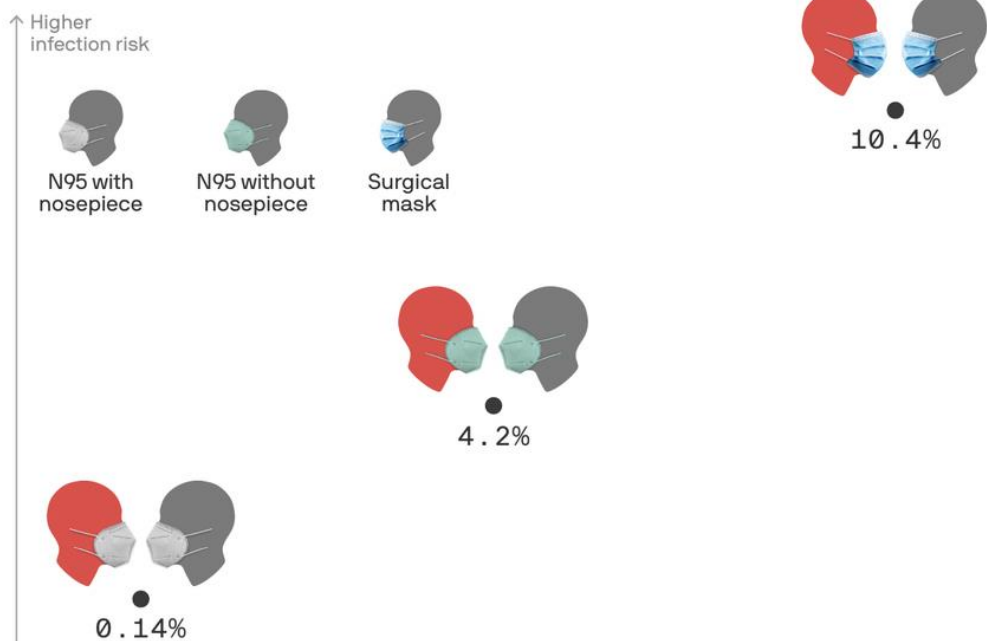
d. A todos los pacientes con COVID19 ya que los ingresados son menos graves y funcionará mejor en ellos

Pregunta 19 Enunciado de la pregunta

En la figura que parece inmediatamente después aparecen los porcentajes de éxitos en 3 tratamientos diferentes para una enfermedad crónica. Elija la opción correcta que se aplicaría a la vista de este gráfico:

Risk of COVID-19 infection between infected and non-infected individuals with different mask combinations

Mean infection risk after 20 minutes of speaking



Fuente: <https://venngage.com/blog/misleading-graphs/>

- a. La gráfica es incorrecta porque los porcentajes que aparecen en ella no suman el 100% como deberían.
- b. La gráfica es correcta porque los porcentajes de éxito en los cuatro tratamientos guardan la relación que guardan las áreas de los rectángulos que representan cada tratamiento.
- c. La gráfica es correcta porque las diferencias entre tratamientos quedan bien reflejadas en ella.
- d. La gráfica es incorrecta pues el eje de abscisas, al no aparecer desde 0 magnifica las diferencias entre tratamientos.

Pregunta 20 Enunciado de la pregunta

En el caso de un test de comparación de dos proporciones de una cola ¿Cuál es la pareja de hipótesis nula e hipótesis alternativa correcta?

- a. $H_0 \equiv p_1 = p_2$, $H_1 \equiv p_1 < p_2$
- b. Ninguna de las otras alternativas es correcta
- c. $H_0 \equiv p_1 > p_2$, $H_1 \equiv p_1 \neq p_2$
- d. $H_0 \equiv p_1 \neq p_2$, $H_1 \equiv p_1 = p_2$

Pregunta 21 Enunciado de la pregunta

Una niña tiene un valor tipificado del peso, a los 6 años, de 1.8. A los 10 años, su valor tipificado del peso es de 2.1. ¿Qué le ha pasado? Señale la opción correcta de entre las cuatro explicaciones que se le ofrecen.

- a. La niña ha aumentado de peso como todos los niños al pasar de los 6 a los 10 años porque la diferencia entre los valores tipificados es pequeña.
- b. La niña ha aumentado de peso frente a su población, ya que a los 10 años está a 2.1 desviaciones típicas de la media de su población mientras que a los 6 estaba a 1.8 desviaciones típicas de la media de la población.
- c. La niña ha aumentado de peso ya que a los 10 años su valor tipificado es mayor que a los 6 años
- d. La niña ha aumentado mucho de peso ya que su valor tipificado a los 10 pasa de 2.

Pregunta 22 Enunciado de la pregunta

Señale la definición correcta.

- a. Un contraste de hipótesis es un procedimiento para confirmar que la hipótesis nula es falsa.
- b. Un contraste de hipótesis es un procedimiento estadístico para decidimos por una de estas dos hipótesis: hipótesis nula e hipótesis alternativa.
- c. Un contraste de hipótesis es un procedimiento estadístico para estimar la diferencia entre la hipótesis nula y la hipótesis alternativa
- d. Un contraste de hipótesis es un procedimiento estadístico para demostrar que la hipótesis que propone el investigador es cierta.

Pregunta 23 Enunciado de la pregunta

Un grupo de 135 pacientes que sufrían de estenosis de columna lumbar y además tenían espondilolistesis degenerativa fueron divididos en dos grupos al azar de 67 y 68 personas cada uno a los que se les sometió a dos intervenciones quirúrgicas: el primer grupo fue sometido a una cirugía de descompresión y una de fusión (CDF), mientras que al segundo se le aplicó una cirugía de fusión (CF). La comparación se hizo para ver, si en promedio, el Índice de Discapacidad Oswestry (ODI), conforme mayor más discapacidad, era diferente en las personas sometidas a esos dos tipos de intervenciones. Los resultados de dicho índice a los 2 años de la intervención fueron:

MuestraCDF $\bar{x}=41.2$, $s=13.3$ MuestraCF $\bar{x}=40.8$, $s=13.2$

A la vista de estos resultados diga si, en promedio, hay diferencias entre las dos medias poblacionales y hágalo de manera fiable suponiendo que la diferencia mínima de interés entre ambas medias sería 5 y que se desearía detectarla cuando realmente exista con una probabilidad de 0.80.

Solución:

(Se presenta ahora un informe que usted debe rellenar. El rellenado se hará unas veces con una elección múltiple, otras con un número y otras con un texto libre; siga las instrucciones sobre el formato de introducción de la información rigurosamente.)

Para emitir el siguiente informe hemos de empezar suponiendo que las variables aleatorias que define el problema

son:

El test que resuelve el problema es

el

Para llevarlo a cabo necesitamos que las variables sean normales. *(Siga con el ejercicio independientemente de su última contestación)*

La instrucción de la librería BioestadísticaR que debería emplear para resolver el problema completo es:

Los resultados de la ejecución correcta figuran en el siguiente recuadro:

```
#-----
```

```
# Información muestral
```

```
Niveles de agrupación: 1, 2
```

```
[1] Para grupo = 1
```

```
  n1 = 67
```

```
  m1 = 41.2
```

$$s_1 = 13.3$$

$$\text{sem}_1 = 1.6249$$

$$95\%-\text{IC}(\mu_1) = (37.9559, 44.4441)$$

[2] Para grupo = 2

$$n_2 = 68$$

$$m_2 = 40.8$$

$$s_2 = 13.2$$

$$\text{sem}_2 = 1.6007$$

$$95\%-\text{IC}(\mu_2) = (37.6049, 43.9951)$$

Test de homogeneidad de varianzas. $F_{\text{exp}} = (\text{Var}_1/\text{var}_2)$

$$F_{\text{exp}} = 1.0152, \text{g.l.}_1 = 66, \text{g.l.}_2 = 67, p = 0.9506$$

Diferencia de medias (grupo[1] - grupo[2])

Diferencia a contrastar: $\mu_0 = 0$

a) Test de Student (varianzas homogéneas)

$$t_{\text{exp}} = 0.1754, \text{g.l.} = 133, p = 0.861$$

$$95\%-\text{IC}(\mu_1 - \mu_2) = (-4.1113, 4.9113)$$

b) Test de Welch (varianzas no homogéneas)

$$t_{\text{exp}} = 0.1754, \text{g.l.} = 132.93, p = 0.8611$$

$$95\%-\text{IC}(\mu_1 - \mu_2) = (-4.1116, 4.9116)$$

Estudio de la potencia: $\delta = 5 \rightarrow [-5, 5]$, potencia $\theta = 80\%$

$$60\%-\text{IC}(\mu_1 - \mu_2) = (-1.5258, 2.3258)$$

---[-(-|-)]---- potencia > 80%

Leyenda: --(---)-- --[---|---]--

IC- IC+ - δ 0 + δ

Estimación del tamaño muestral para detectar una diferencia $\delta=5$ con potencia $\theta=80\%$

(1) Considerando las varianzas homogéneas:

$(n_1 = n_2) \geq 112$ casos en cada grupo

(2) Considerando las varianzas heterogéneas: $k=s_1/s_2=1.0076$, $(g.l.'=132.98)$

$n_1 \geq 113$ casos en el grupo [1]

$n_2 \geq 112$ casos en el grupo [2]

A partir de estos datos la cantidad experimental del test es (redondeada a 2 decimales) $t_{exp} = \text{[]}$, g.l.(redondeados si necesario, a un decimal): [] , P(redondeado a tres decimales): [] .

Con estos resultados podemos afirmar:

. (siga, independientemente del resultado que haya reportado antes.)

Para ver si esta decisión es fiable, se ha calculado un intervalo para la diferencia de dos medias a un error del (debe poner un porcentaje sin decimales) [] %, que nos ha dado los valores (redondeado a dos decimales) $(\text{[]}; \text{[]})$ y como la diferencia mínima que se desea detectar(un decimal) [] está de dicho intervalo para detectarla y no lo hemos hecho con lo que la conclusión fiable.

Pregunta 24 Enunciado de la pregunta

En los pacientes críticos se sospecha que la presión arterial sistólica (PAS) medida en ambos brazos no es la misma. Con objeto de verificar tal hipótesis, se tomó una muestra de pacientes críticos a los que se les midió la presión sistólica a la entrada en la UCI en ambos brazos, obteniéndose los datos que figuran en la siguiente tabla (que pueden considerarse como provenientes de variables aleatorias normales):

Paciente	1	2	3	4	5	6	7	8	9	10
PAS brazo derecho (pasd)	152	183	177	160	146	200	154	133	119	173
PAS brazo izquierdo (pasi)	169	192	167	168	129	202	160	122	164	174

- a) ¿Qué se puede decir de la sospecha inicial?
- b) Independientemente de lo que haya obtenido en el apartado anterior, suponga que la sospecha inicial no se ha corroborado. ¿Deberíamos ampliar la muestra si, al realizar el test a un nivel de error del 5%, deseáramos detectar una diferencia de 7 mmHg entre ambos brazos con una probabilidad del 80%?

Solución:

(Se presenta ahora un informe que usted debe rellenar. El rellenado se hará unas veces con una elección múltiple, otras con un número y otras con un texto libre; siga las instrucciones sobre el formato de introducción de la información rigurosamente.)

Apartado a)

Para emitir el siguiente informe hemos de empezar suponiendo que las variables aleatorias que define el problema son:

El test que resuelve el problema es

el

Según se nos informa en el enunciado las variables .

(Siga con el ejercicio independientemente de su última contestación)

Suponiendo que los datos de la PAS en el brazo derecho se han guardado en una variable denominada pasd y los del brazo izquierdo en una variable denominada pasi, la instrucción de la librería BioestadísticaR que debería emplear para resolver el problema completo es:

Los resultados de la ejecución correcta figuran en el siguiente recuadro:

```
# -----  
# Información muestral  
[1] Variable pre: pasd  
  n1 = 10  
  m1 = 159.7  
  s1 = 24.2489  
  sem1 = 7.6682  
  95%-IC( $\mu_1$ ) = (142.3534, 177.0466)  
  
[2] Variable post: pasi  
  n2 = 10  
  m2 = 164.7  
  s2 = 24.4452  
  sem2 = 7.7302  
  95%-IC( $\mu_2$ ) = (147.213, 182.187)
```

```
# Correlación (pre, post): r(pasd,pasi)
r = 0.7414

# Normalidad de la diferencia (Test de Shapiro-Wilk)
W =0.9012, g.l. = 10, p = 0.226

# Diferencia pre-post;pasd-pasi
n = 10
m = -5
s = 17.5119
sem = 5.5377
95%-IC(μ)= (-17.5273, 7.5273)

# Valor de la diferencia media a contrastar: μ₀ = 0
texp =0.9029, g.l. =9, p = 0.3901
95%-IC(μ₁-μ₂) = (-17.5273, 7.5273)

# Estudio de la potencia: δ = 7 -> [-7, 7], potencia θ= 80%
60%-IC(μ₁-μ₂) = (-9.8921, -0.1079)
--(-)|---]--- potencia < 80%

Leyenda: --(---)-- --[---|---]--
          IC- IC+   -δ  0 +δ
```

A partir de estos datos la cantidad experimental del test es (redondeada a 2 decimales) $t_{exp} = \square$, g.l. (redondeados si necesario, a un decimal): \square , P(redondeado a tres decimales): \square .

Con estos resultados podemos afirmar:

. (siga, independientemente del resultado que haya reportado antes.)

Aparatado b)

Para ver si esta decisión es fiable, se ha calculado un intervalo para la diferencia de dos medias a un error del (debe poner un porcentaje sin decimales) \square %, que nos ha dado los valores (redondeado a dos decimales) (\square ; \square) y como la diferencia mínima

que se desea detectar(dos decimales) está de dicho intervalo para detectarla y no lo hemos hecho con lo que la conclusión .

Pregunta 25 Enunciado de la pregunta

¿En un contraste de hipótesis una decisión por H_0 es siempre fiable? ¿Por qué? Elija la opción correcta

- a. Siempre es fiable porque el error β lo elige el investigador pequeño.
- b. Nunca es fiable porque la hipótesis nula no es nunca cierta
- c. No es siempre fiable a menos que se haya calculado, a priori, el tamaño de muestra para detectar una alternativa de interés con una potencia que le parezca al investigador suficientemente grande.
- d. No siempre es fiable ya que el error α es en ese caso grande.

Pregunta 26 Enunciado de la pregunta

Una persona (A) a los 25 años de edad tiene el nivel de fibrinógeno en plasma en el percentil 50. Su hijo(B), a la misma edad, está en el percentil 80 del nivel de fibrinógeno en plasma. Señale que frase expresaría mejor lo que le ha pasado a B con respecto a A.

- a. B ha mejorado sensiblemente su nivel de colesterol con respecto a A por haber subido en el percentil para su población.
- b. B ha empeorado en su nivel de fibrinógeno en plasma con respecto a A y como ambos comparten una importante carga genética es posible que ese incremento se deba a factores diferentes que la genética.
- c. No son comparables ambos niveles puesto que A y B pertenecen a poblaciones diferentes.
- d. B ha empeorado con respecto a A ya que sólo el 50% de su población tiene un valor más alto que el suyo cuando en el caso del A es un 20%

Pregunta 27 Enunciado de la pregunta

¿Qué le pasa a la potencia de un test cuando varía el error de Tipo I?

- a. Cuando se incrementa el error de Tipo I la potencia disminuye
- b. Cuando se disminuye el error de Tipo I la potencia aumenta
- c. La potencia y el error de Tipo I no están relacionados

- d. Cuando se incrementa el error de Tipo I la potencia se hace mayor

Pregunta 28 Enunciado de la pregunta

Dos estudios el A y el B, estiman con el mismo tamaño de muestra la efectividad de una vacuna, siendo sus resultados los que figuran en la siguiente tabla. Señale la razón correcta de las diferencias entre ambos estudios

Estudio	Estimación Puntual Efectividad	IC Efectividad
A	61.3%	(53.5%; 69.0%)
B	61.3%	(55.5%;67.0%)

- a. El estudio B da un intervalo más estrecho debido a una mayor confianza.
- b. Ambos estudios dan la misma estimación puntual luego la diferencia en los intervalos será debida al tamaño de muestra.
- c. El estudio A da un intervalo más estrecho debido a una menor confianza.
- d. El estudio B da un intervalo más estrecho debido a una menor confianza.

Pregunta 29 Enunciado de la pregunta

¿Cómo son las categorías de carácter con el que se ha formado la siguiente tabla?

Grupo sanguíneo	A	B	0	Total
Frecuencia	65	25	150	300

Señale la opción correcta.

- a. No son ni excluyentes ni exhaustivas
- b. Son exhaustivas, pero no excluyentes
- c. Son excluyentes, pero no exhaustivas.
- d. Son excluyentes y exhaustivas

Pregunta 30 Enunciado de la pregunta

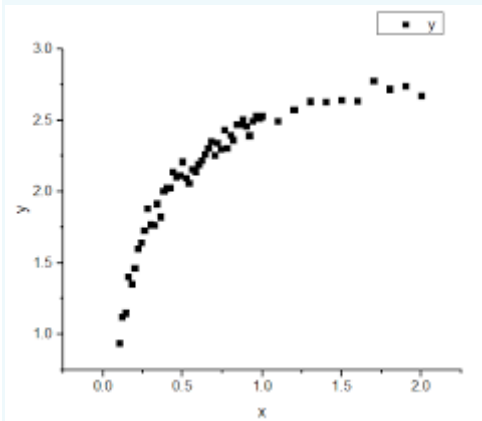
En una muestra de 26 ingresados en el hospital por COVID19 se encontró que 10 sobrevivieron durante esa hospitalización. Esas 26 personas eran de un subgrupo de muy alto riesgo. La estimación puntual de la probabilidad de morir durante la estancia fue de $p^{\wedge}=0.43$, siendo el intervalo al 95% (0.24;0.63) (los resultados han sido

redondeados para facilitar las operaciones sin calculadora). Suponiendo que uno desea tener una precisión de 0.19 en la estimación de dicha proporción qué debería hacerse frente a estos resultados.

- a. La anchura del intervalo es 0.39 que es una anchura inferior a la que se pide por lo que habría que disminuir el tamaño de la muestra.
- b. La anchura del intervalo es 0.39 que es una anchura superior a la que se pide por lo que habría que disminuir el tamaño de la muestra.
- c. La anchura del intervalo es 0.39 que es una anchura inferior a la que se pide por lo que habría que aumentar el tamaño de la muestra.
- d. La anchura del intervalo es 0.39 que es una anchura superior a la que se pide por lo que habría que aumentar el tamaño de la muestra.

Pregunta 31 Enunciado de la pregunta

A la vista de la gráfica siguiente ¿podríamos ajustar el modelo de regresión lineal simple?. Elija la alternativa correcta.



- a. Sí se puede porque como se ve en la gráfica se verifican todas las hipótesis del modelo de regresión lineal.
- b. Sí, porque se verifica la hipótesis de normalidad.
- c. No, porque no se verifica la hipótesis de homogeneidad de varianzas.
- d. No, porque no se verifica la hipótesis de linealidad.

Pregunta 32 Enunciado de la pregunta

Señale la pareja de valores, riesgo relativo (R) y odds ratio (O), que sea coherente, de entre las 4 que se presentan.

- a. $O=0.25$ $R=1$

- b. $O=1$ $R=0.18$
- c. $O=1/4$ $R=4$
- d. $O=0.25$ $R=0.18$

Pregunta 33 Enunciado de la pregunta

Un biomarcador clásico de la enfermedad de Alzheimer es el nivel de proteína beta-amiloide en el líquido cefalorraquídeo, que se analizan a partir del líquido extraído mediante una punción lumbar. Se expresa como concentración en pg/ml. Dígase el tipo de variable aleatoria que es dicho biomarcador.

- a. Es continua porque entre dos concentraciones de tal biomarcador siempre existe otra concentración posible
- b. Es discreta porque entre dos niveles del biomarcador no se puede encontrar siempre otro nivel del biomarcador
- c. Es discreta porque no se puede expresar ese biomarcador con infinitos decimales
- d. Es continua porque la distribución de todo biomarcador es simétrica

Pregunta 34 Enunciado de la pregunta

Señale el estadístico de contraste del test de Wilcoxon para muestras apareadas y diga qué valores del mismo nos llevarían a rechazar la hipótesis nula. (Test de dos colas)

- a. Estadístico de Contraste: Suma de los rangos de todas las diferencias: Valores: Altos o Bajos
- b. Estadístico de Contraste: Suma de los rangos de todas las diferencias: Valores: Altos
- c. Estadístico de Contraste: Suma de los rangos de todas las diferencias: Valores: Bajos
- d. Estadístico de Contraste: Suma de los rangos de las diferencias negativas. Valores: Altos o Bajos

Pregunta 35 Enunciado de la pregunta

¿En un contraste de hipótesis una decisión por H_1 es fiable? ¿Por qué? Elija la opción correcta

- a. No es fiable porque nadie nos asegura que el intervalo para el parámetro objeto del estudio no contenga en su interior al verdadero valor del parámetro
- b. Si es fiable porque al ser significativo el error de Tipo II es pequeño.

- c. No es fiable porque para ello el intervalo para el verdadero valor del parámetro propuesto por la hipótesis nula debería abarcar en su interior a dicho valor.
- d. Si es fiable, porque salvo el error de Tipo I aparejado a esa decisión (cuya probabilidad es tan pequeña como haya deseado el investigador) la decisión es correcta.

Pregunta 36 Enunciado de la pregunta

¿Qué mide el estadístico χ^2 ? ¿Qué valores dan lugar al rechazo de la hipótesis nula?

- a. La discrepancia entre los valores observados y los valores esperados. Discrepancias grandes dan lugar a valores pequeños y por tanto al rechazo de H_0 .
- b. La discrepancia entre los valores observados y los valores esperados supuesta cierta la hipótesis nula. Discrepancias pequeñas dan lugar a valores grandes y por tanto al rechazo de H_0 .
- c. La discrepancia entre los valores observados supuesta cierta H_0 y los valores esperados. Discrepancias pequeñas dan lugar a valores pequeños y por tanto al rechazo de H_0 .
- d. La discrepancia entre los valores observados y los valores esperados supuesta cierta la hipótesis nula. Discrepancias grandes dan lugar a valores grandes y por tanto al rechazo de H_0 .

Pregunta 37 Enunciado de la pregunta

La adherencia a un programa de abandono del hábito tabáquico es determinante para el éxito del mismo. Una forma de medir la adherencia es la participación en tres actividades de dicho programa que se sitúan en las primeras fases del mismo. Conforme ese número es mayor, mayor será la adherencia al tratamiento.

Nº Asistencias(x)	0	1	2	3
%	11%	19%	50%	20%

¿Cuál será el punto de corte para decidir una asistencia baja con un error del 5%? (Elija la opción correcta)

- a. Será el valor 0 puesto que $P(x \leq 0) = 0.11$ que aun no siendo el 5% es el primero en el que se desborda esa cantidad
- b. No se puede calcular ese límite porque no hay ningún valor en el que la probabilidad de que x sea menor o igual que él valga exactamente 0.05

- c. Será el 2, puesto que en él la probabilidad de que x sea menor o igual que él es 0.30 que es francamente mayor que 0.05
- d. Será el 1 puesto que $P(x \leq 1) = 0.30$ que, aunque mayor que 0.05 es muy cercana a él

Pregunta 38 Enunciado de la pregunta

¿Cuándo es preferido un test paramétrico a uno no paramétrico para hacer el test de comparación de dos medias muestras apareadas?

- a. Cuando las variables aleatorias no sean normales o el número de parejas sea inferior a 60.
- b. Cuando las variables aleatorias sean normales o el número de parejas sea superior a 60.
- c. Cuando las variables aleatorias no sean normales y el número de parejas sea inferior a 60.
- d. Siempre.

Pregunta 39 Enunciado de la pregunta

La ecuación que liga a la edad (en años) con la talla(cm) en una muestra de 15000 niños entre 6 y 16 años, siendo y =Talla y la X =edad en años, es:

$$y = 83.0 + 5.1 \cdot x$$

Elija la interpretación correcta

- a. La altura en el origen indica que la talla media a los 6 años sería de 83.0cm. La pendiente indica que por cada año que aumenta la edad, en promedio la talla sube 5.1cm
- b. La altura en el origen indica que la talla media a los 0 años sería de 83.0cm. La pendiente indica que por cada año que aumenta la edad, la talla de un niño sube 5.1cm
- c. La altura en el origen indica que la talla media a los 0 años sería de 83.0cm. La pendiente indica que por cada año que aumenta la edad, en promedio la talla sube 5.1cm
- d. La altura en el origen indica que la talla media a los 6 años sería de 83.0cm. La pendiente indica que por cada año que aumenta la edad, la talla de un niño sube 5.1cm

Pregunta 40 Enunciado de la pregunta

En un estudio retrospectivo de una enfermedad no rara, ¿se puede calcular el riesgo relativo? ¿Porqué? Señale la opción correcta.

- a. Sí, porque en un retrospectivo se pueden calcular todas las medidas.
- b. Sí, porque se puede calcular cualquier probabilidad
- c. No, porque no se puede calcular la probabilidad de que teniendo la enfermedad se tenga el factor de riesgo.
- d. No, porque no se puede calcular la probabilidad de tener la enfermedad cuando se tiene el factor de riesgo.

Pregunta 41 Enunciado de la pregunta

El test chi-cuadrado no se debe hacer si no se cumplen las condiciones de validez. Esas condiciones de validez previenen de que los valores esperados supuesta cierta la hipótesis nula sean muy pequeños. Elija para esa razón la explicación correcta de entre las cuatro.

- a. Los valores esperados no deben ser muy pequeños ya que al estar en el numerador de la chi-cuadrado, ésta podría ser pequeña por el valor del numerador y no porque las discrepancias sean grandes
- b. Los valores esperados no deben ser muy pequeños ya que al estar en el denominador de la chi-cuadrado, ésta podría ser pequeña por el valor del denominador y no porque las discrepancias sean grandes
- c. Los valores esperados no deben ser muy grandes ya que al estar en el denominador de la chi-cuadrado, ésta podría ser pequeña por el valor del denominador y no porque las discrepancias sean grandes
- d. Los valores esperados no deben ser muy pequeños ya que al estar en el denominador de la chi-cuadrado, ésta podría ser grande por el valor del denominador y no porque las discrepancias sean grandes

Pregunta 42 Enunciado de la pregunta

A partir de los dos esquemas de muestreo siguientes, A y B, señale el tipo de muestras.

Esquema A: Para comparar la proporción de reacciones adversas a dos fármacos se toma una muestra y se divide al azar en dos, administrándosele a una un fármaco y a la otra el otro fármaco.

Esquema B: Para comparar la proporción de reacciones adversas a dos fármacos se administran ambos de manera consecutiva a una muestra de 100 individuos, recogiendo tras la administración si se presentaba o no la reacción adversa.

- a. Esquema A: Apareadas, Esquema B: Independientes.
- b. Esquema A: Independientes, Esquema B: Apareadas
- c. Esquema A: Apareadas, Esquema B: Apareadas
- d. Esquema A: Independientes, Esquema B: Independientes

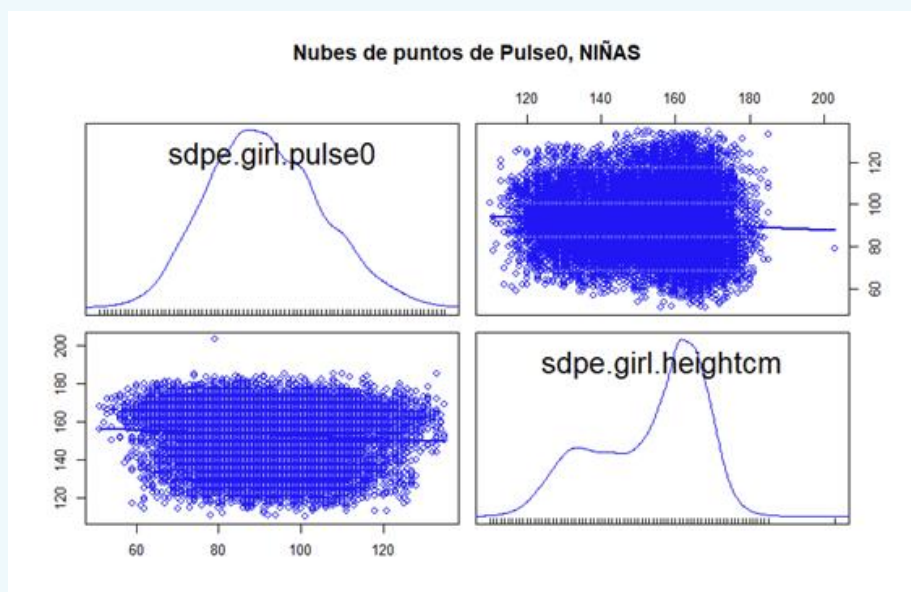
Pregunta 43 Enunciado de la pregunta

En el archivo `sdpe.RData` aparecen los valores del pulso en una muestra de niños desde 6 a 16 años a distintos tiempos tras una carrera de 1000m o 2000m según la edad. Las variables se denominan `pulse0`, `pulse1`, `pulse5` y `pulse10`, haciendo referencia al instante del inicio de la carrera (0'), al minuto (1'), a los 5 minutos (5') y a los 10 minutos (10'). El archivo está dividido en dos uno `sdpe.boy`(para niños) y otro `sdpe.girl`(para niñas). Se desea estudiar si existe asociación entre la talla de las niñas (`sdpe.girl$height`) y el pulso en el tiempo 10 y como es la ecuación que liga a la talla con el pulso en el instante 10. Además, se desea interpretar los coeficientes de la recta y dar una medida de la fuerza de la asociación entre ambas variables.

Solución:

(Se presenta ahora un informe que usted debe rellenar. El rellenado se hará unas veces con una elección múltiple, otras con un número y otras con un texto libre; siga las instrucciones sobre el formato de introducción de la información rigurosamente.) En primer lugar, hemos de pedir una nube de puntos que nos permita verificar, al menos, algunas hipótesis del modelo de regresión lineal simple. Hecho esto se obtiene la matriz de nubes de puntos siguiente:

Esta matriz de nubes de puntos



Esa matriz de nubes de puntos presenta dos nubes de puntos diferentes, la primera la del pulso en el instante 1(pulse10) sobre la talla (heightcm) y la segunda la de la talla sobre el pulso en el instante 10. Para chequear las hipótesis del modelo de regresión lineal nos quedaremos con la nube de:

Independientemente de lo contestado observe la gráfica de la primera fila, segunda columna. En ella se puede ver que la relación entre la talla y el pulso no

será:

Con respecto a la hipótesis de normalidad de la variable y para cada valor de x,

podemos decir:

Con respecto a la hipótesis de homogeneidad de varianzas de la variable y para cada valor de x, podemos

decir:

Suponga de todas maneras que se verifican las hipótesis del modelo de regresión lineal.

A partir de ahí calcularemos la recta de mejor ajuste a la nube de puntos con el comando (note que el símbolo ~ se representa con

\~):

```
Call:
lm(formula = sdpe.girl$pulse1 ~ sdpe.girl$heightcm)

Residuals:
    Min       1Q   Median       3Q      Max
-66.335 -12.533  -0.591  12.879  55.723

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.139e+02  1.332e+00   85.53  <2e-16 ***
sdpe.girl$heightcm 1.372e-01  8.661e-03   15.84  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 18.39 on 21017 degrees of freedom
Multiple R-squared:  0.0118, Adjusted R-squared:  0.01175
F-statistic: 251 on 1 and 21017 DF, p-value: < 2.2e-16

Pearson's product-moment correlation

data: sdpe.girl$pulse5 and sdpe.girl$heightcm
t = -1.6204, df = 21017, p-value = 0.1052
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.024691896  0.002342853
sample estimates:
      cor
-0.01117656
```

Los resultados figuran en la tabla anterior. De ella se puede decir que para la pendiente la estimación puntual es (máximo dos decimales redondeados): que es: . Ello debido a que el test que tiene de hipótesis nula: con un P (tres decimales, poniendo 0.001 si $P < 0.001$): . La interpretación de la pendiente es la siguiente:

No comentaremos la altura en el origen (Intercept) porque no están dispuestos los datos en forma de una interpretación biológica relevante. Veamos cómo podemos estimar la fuerza de la asociación entre las dos variables pulse10 y heightcm. Calcularemos el coeficiente de correlación de Pearson con la instrucción:

El valor del coeficiente de correlación de Pearson es (dos decimales redondeados): que nos dice que ese coeficiente es: .

El cuadrado de r y después multiplicado por cien nos indica que el porcentaje de variabilidad del pulse10 que depende de la talla (heightcm) en niñas es de un(dos decimales en porcentajes): %

Tras todo ello y a la vista de los resultados:

Pregunta 44 Enunciado de la pregunta

El número de nuevos diagnósticos anuales de acondroplasia en una población de medio tamaño (250,000 habitantes) sigue una distribución: (elija la opción correcta).

- a. Sigue una Poisson ya que el suceso de dar un diagnóstico de acondroplasia es raro y la población a la que se aplica es grande.
- b. Sigue una Poisson ya que es un número que va de 0 a infinito.
- c. Sigue una Normal porque el número de casos de acondroplasia es una continua y es simétrica.
- d. Sigue una Binomial con n grande y p grande porque el tamaño de la población es grande.