

1. Práctica 1: biología sanitaria

1.1. Estadística descriptiva con Calc

La hoja 1 de este [fichero](#) contiene es parte del famoso estudio sobre enfermedades cardiacas realizado en Framingham. Hay algunos datos de las variables `sex1` (género), `age1` (edad) y `systbp1` (presión sanguínea sistólica).

Antes de lanzarse a hacer estadística, hay que comprobar si hay registros (filas) incompletas y, en caso de que los haya, decide qué hacer con ellos. Si no tienes ninguna buena razón para hacer lo contrario, lo mejor es eliminarlos (a no ser que los datos ausentes afecten a una variable que no vasa estudiar). [Aquí](#) está incrustado el fichero Calc con los cálculos hechos (recuerda que con casi todos los navegadores de internet necesitarás descargar este pdf a tu disco duro para poder acceder a ficheros incrustados).

1. Estudia la distribución de la muestra en cuanto a la variable género.
Se trata de calcular la tabla de frecuencias (valores discretos repetidos). Para el recuento, hay que usar las funciones `=CONTAR.SI(a:2;a:120; "Male")` y `=CONTAR.SI(a:2;a:120; "Female")` porque estamos contando palabras y no números (ahí se usa FRECUENCIA)
2. Representa gráficamente dicha información.
Diagramas de sectores y de barras, mejor el segundo. Si usas el primero, asegúrate de que sabes rotular los sectores con las frecuencias y el tanto por cien correspondiente.
3. Calcula la tabla de frecuencias para la variable `age1`.
En la medida de lo posible, usa clases de la misma amplitud. En este caso, de amplitud 4,
4. Calcula la edad media de dos formas diferentes: con la función SUMA y a partir de la tabla de frecuencias absolutas (decide tú cuantas clases haces). ¿Valen lo mismo? ¿A qué se debe la discrepancia?
En un caso el cálculo es aproximado (tabla de frecuencias) y en el otro (función SUMA) exacto.
5. Calcula de forma aproximada (es decir, partir de la tabla de frecuencias relativas acumuladas) los cuartiles y el percentil 40 de la variable edad.
6. Calcula con la correspondiente función de Calc, los cuartiles y el percentil 40 de la variable edad.
Puedes resolverlo todo con una única función: por ejemplo, `CUARTIL(datos; 1)` es lo mismo que `PERCENTIL(datos; 0,25)`
7. Respecto de la variable edad, ¿en qué percentil se encuentra un individuo de 54 años?
Si lo haces a partir de la tablade frecuencias acumuladas, hay que buscar la clase en la que se encuentra dicho individuo.
8. Calcula la varianza poblacional de la variable presión arterial de las mujeres usando la tabla de frecuencias y con la función de Calc.
Si ordenas la primera columna alfabéticamente (selección extendida) podrás separar fácilmente la subpoblación.

9. Compara la distribución de presión sanguínea por sexo. ¿En qué género (sexo) hay más dispersión?

Una vez que has ordenado las variables por sexo, usa directamente la función VARP. Como las medias que se obtienen son semejantes, puedes comparar directamente las varianzas. Si lo prefieres, puedes calcular los correspondientes coeficientes de variación, para ello, usa la función DESVEST (que proporciona la desviación típica muestral). Mientras que las medias son similares entre hombres y mujeres, se aprecia que la variabilidad es sensiblemente mayor en las mujeres. Una pregunta importante es si esa diferencia es significativa (y extensible a toda la población de la que proviene la muestra) o si se puede achacar al azar, es decir, una coincidencia que tiene que ver con la muestra concreta con la que trabajamos. Para responder a esa pregunta tendremos que aprender algo de probabilidad.

10. La varianza poblacional de muestra de 10 datos vale 3. ¿Cuánto vale la varianza muestral de dicha muestra?

1.2. Cálculo de la recta de regresión con Calc

La hoja 2 del fichero contiene los valores de las presiones sistólica y diastólica

1. ¿Existe alguna relación entre ellas?

Al representar el diagrama de dispersión se aprecia cierta relación lineal. Hay algunos datos atípicos, y hay que decidir qué hacer con ellos; ¿son plausibles? el habitual analizar su influencia calculando el coeficiente de correlación con y sin ellos.

La ecuación de la recta, con 4 cifras significativas, es

$$pres - diast = 0.4065 * pres - sist + 28.28)$$

2. ¿Qué presión diastólica se espera que tenga, según el modelo lineal, un individuo con una presión sistólica de 120?

Hay que sustituir 120 en $pres - diast = 0.4065 * pres - sist + 28.28$.

3. Calcula los residuos para la variable presión diastólica.

Observado menos predicho (mira la fórmula usada en la hoja de cálculo con las soluciones).

4. Si distinguimos los datos por género, ¿cuál se ajusta mejor al modelo lineal?

Con los datos aportados originalmente no podías hacer el cálculo, faltaba la variable sexo. En las soluciones la hemos incluido; ordenado los datos de acuerdo a dicha variable, y calculado las rectas de regresión. La que tiene mayor coeficiente de correlación lineal es, en principio, la que mejor se ajusta (ajuste ligeramente superior para hombres: $r = 0.7883$ frente a 0.7691).