Seguimos avanzando con los conceptos de estadística descriptiva.

Ahora vas a practicar con las tablas de contingencias.

Para entenderlas es mejor ir paso a paso practicando.

# ABRE LA TABLA DE DATOS “esoph”

Si has hecho la primera hoja de trabajo, ahora sólo tendrás que copiar la siguiente tabla:

*Más información: https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/esoph.html*

|  |  |  |  |
| --- | --- | --- | --- |
| Nombre Variable | ¿Qué característica mide? | ¿Qué unidades tiene? | ¿Qué tipo de variable es? |
| agegp | Agrupa por años de edad | Años | Categorica |
| alcgp | Agrupoa por consumo de alcohol diario | mg/dia | Categorica |
| tobgp | Agrupa por consumo de tabaco diario | Mg/dia | Categorica |
| ncases | Cantidad de Casos | Cantidad de caso | Cuantitativa |
| ncontrols | Numero de Control | NA | Cuantitativa |

# TABLAS DE FRECUENCIAS y PROPORCIONES

Calcula la tabla de contingencias para la variable cualitativa: “agegp”

|  |  |  |  |
| --- | --- | --- | --- |
| Grupos en años | Frecuencia | Frec. Relativa | Proporción (porcentaje) |
| 25-34 | 15 | 0.170454545 | 17.05% |
| 35-44 | 15 | 0.170454545 | 17.05% |
| 45-54 | 16 | 0.181818182 | 18.18% |
| 55-64 | 16 | 0.181818182 | 18.18% |
| 65-74 | 15 | 0.170454545 | 17.05% |
| 75+ | 11 | 0.125 | 12.50% |
| Total | 88 | 1 | 100.00% |

# DIAGRAMA DE BARRAS y SECTORES

Describe esta tabla de datos con un diagrama de sectores y de barras.

|  |
| --- |
| Copia el diagrama de barras y el de sectores para representar la tabla anterior. Puedes hacerlo en frecuencia o en proporciones. |

# TABLA DE CONTINGENCIAS

Ahora vamos a observar las parejas de grupos. Entre los grupos de edad (“agegp”) y los grupos de casos (“ncases”).

De esta forma sabremos la distribución de casos de Cáncer por grupos de edad:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Grupos de años | Nº casos grupos | | | | | | | | |
| **1** | **2** | **3** | **4** | **5** | **6** | **8** | **9** | **17** |
| 25-34 | 1 |  |  |  |  |  |  |  |  |
| 35-44 | 2 | 2 | 1 |  |  |  |  |  |  |
| 45-54 | 2 | 2 | 2 | 3 | 2 | 2 |  |  |  |
| 55-64 |  | 2 | 4 | 3 | 2 | 2 | 1 | 2 |  |
| 65-74 | 4 | 2 | 2 | 2 | 2 | 1 |  |  | 1 |
| 75+ | 7 | 3 |  |  |  |  |  |  |  |

Observa la tabla y calcula con Excel el número de pacientes con cáncer por cada grupo de edad. (pista: si tenemos 7 casos de 1 y 10 de 6, el número de pacientes es 7\*1+10\*6=76)

|  |  |
| --- | --- |
| Grupos en años | Pacientes totales |
| 25-34 | 1 |
| 35-44 | 9 |
| 45-54 | 46 |
| 55-64 | 76 |
| 65-74 | 55 |
| 75+ | 13 |
| Total | 200 |

# ABRE LA TABLA DE DATOS “WorldPhones”

Esta tabla de datos es en sí una tabla de contingencias.

Tenemos dos variables:

* Los **años:** ordinal (en filas)
* Los **continentes**: nominal (en columnas)

# CALCULA LAS FRECUENCIAS CONDICIONADAS

Ahora se trata de trabajar el concepto de frecuencias condicionadas e interpretar su significado.

Recuerda que las frecuencias condicionadas son frecuencias relativas condicionando a los grupos (dividir la frecuencia absoluta por el total de observaciones del grupo)

## Frecuencias Condicionadas por columnas (continentes)

Calcula las frecuencias condicionadas por columnas.

En este caso a las columnas tenemos la variable “continentes”.

Para cada continente calcularemos el porcentaje condicionado para los diferentes años.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Year** | **N.Amer** | **Europe** | **Asia** | **S.Amer** | **Oceania** | **Africa** | **Mid.Amer** |
| **1951** | 10% | 9% | 7% | 9% | 9% | 1% | 9% |
| **1956** | 13% | 12% | 11% | 13% | 13% | 14% | 12% |
| **1957** | 14% | 14% | 12% | 14% | 14% | 15% | 13% |
| **1958** | 15% | 15% | 15% | 15% | 15% | 16% | 14% |
| **1959** | 15% | 16% | 16% | 15% | 16% | 17% | 15% |
| **1960** | 16% | 17% | 19% | 16% | 17% | 18% | 17% |
| **1961** | 17% | 18% | 21% | 17% | 18% | 19% | 18% |
| **TOTAL** | 10% | 9% | 7% | 9% | 9% | 1% | 9% |

## Dibuja el diagrama de barras o sectores para el continente europeo

|  |
| --- |
| Dibuja el diagrama de barras para el continente europeo e interpreta los resultados. ¿Tiene algún sentido?    Se observa un crecimiento constante y sostenido por año. Seis años después de 1952 el No de Telefonos se habían duplicado vs 1951.  La verdad me parece que sirve para tener una idea del crecimiento y tendencia de la distribución por año |

## Frecuencias Condicionadas por filas (años)

Calcula las frecuencias condicionadas por filas.

En este caso en las filas tenemos la variable “años”.

Para cada año calcularemos el porcentaje de cada continente. (Yo creo que tiene más sentido este cálculo en filas que en columnas, ¿verdad?)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Year | N.Amer | Europe | Asia | S. Amer | Oceania | Africa | Mid. Amer | TOTAL |
| 1951 | 61.67% | 28.96% | 3.86% | 2.44% | 2.21% | 0.12% | 0.75% | 100% |
| 1956 | 59.12% | 29.34% | 4.61% | 2.51% | 2.32% | 1.38% | 0.72% | 100% |
| 1957 | 58.84% | 29.55% | 4.75% | 2.45% | 2.30% | 1.41% | 0.70% | 100% |
| 1958 | 57.84% | 29.75% | 5.63% | 2.40% | 2.27% | 1.40% | 0.71% | 100% |
| 1959 | 57.53% | 30.13% | 5.49% | 2.40% | 2.30% | 1.42% | 0.73% | 100% |
| 1960 | 56.87% | 30.17% | 6.15% | 2.35% | 2.28% | 1.42% | 0.75% | 100% |
| 1961 | 56.34% | 30.47% | 6.39% | 2.36% | 2.28% | 1.41% | 0.76% | 100% |
|  |  |  |  |  |  |  |  |  |

En efecto se puede observar claramente que continente tiene mas o menos representación o proporción por años

## Dibuja el diagrama de barras o sectores para el año 1951 y el 1961. Visualiza el cambio de proporciones.

|  |
| --- |
| Dibuja los dos diagramas de barras o sectores e interpreta los resultados.  Se puede observar que en ambos años America y Europa son por lejos los que mas cantidad de teléfonos de todos los continentes. Sin embargo para 1961 America decrecio y Europa aumento su participación asi como Asia de manera significativa |

¡Estás haciendo un gran trabajo para entender los conceptos a base de ejercicios prácticos!

¡Bien!

¡Vamos a seguir avanzando!