TABLAS DE CONTINGENCIAS

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	Edad del paciente	Años (Categorizados)	Cualitativa Ordinal
alcgp	Consumo de alcohol del paciente	1-4 según gms/día consumidos	Cualitativa Ordinal
tobgp	Consumo de tabaco del paciente	1-6 según gms/día consumidos	Cualitativa Ordinal
ncases	Número de casos	Número (minimo 0-máximo 17), entiendo que es un recuento	Cuantitativa Discreta
ncontrols	Edad del paciente	Años (Categorizados)	Cualitativa Ordinal

TABLAS DE FRECUENCIAS y PROPORCIONES

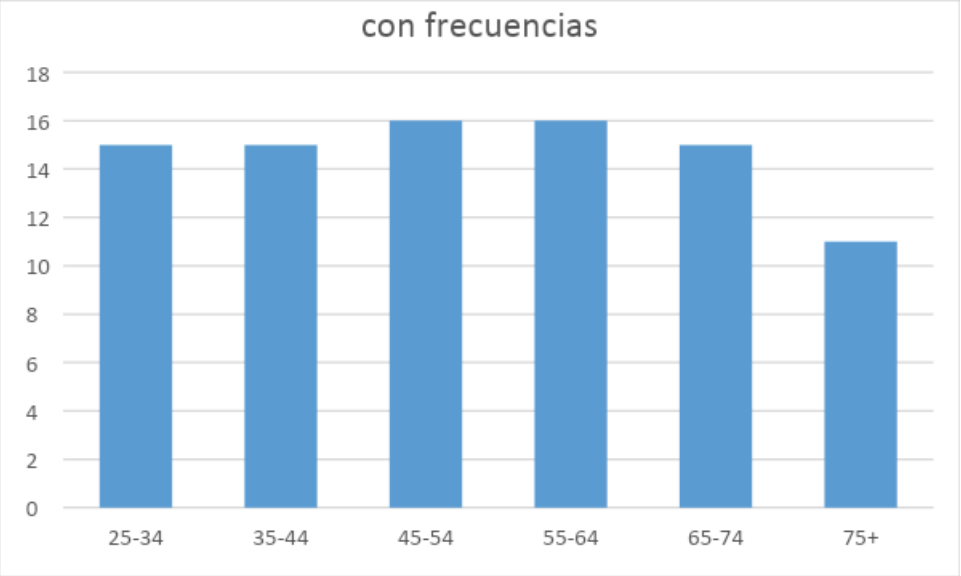
Calcula la tabla de contingencias para la variable cualitativa: “agegp”

Grupos en años	Frecuencia	Frec. Relativa	Proporción (porcentaje)
25-34	15	0,1705	17,05%
35-44	15	0,1705	17,05%
45-54	16	0,1818	18,18%
55-64	16	0,1818	18,18%
65-74	15	0,1705	17,05%
75+	11	0,1250	12,50%
Total	88	1	100%

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.



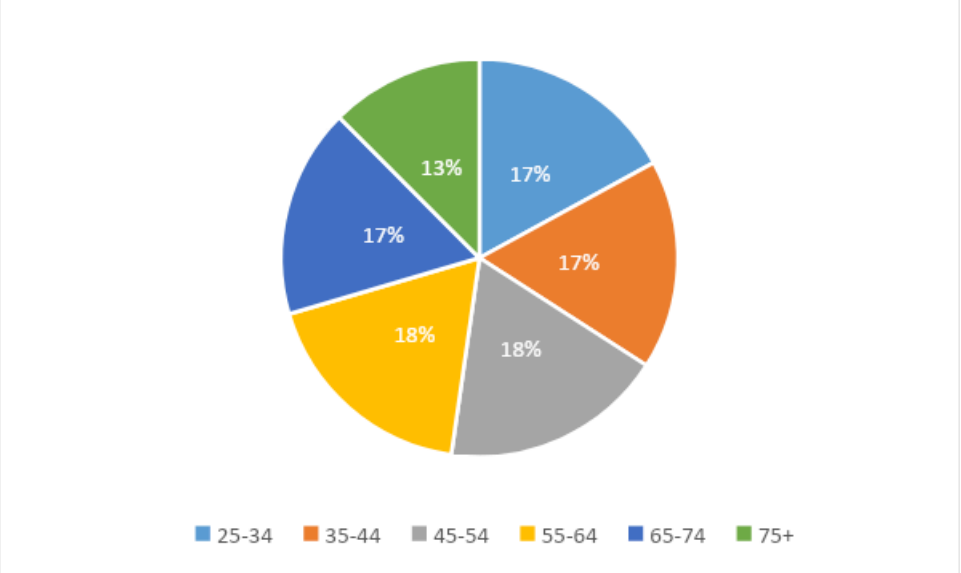


TABLA DE CONTINGENCIAS

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	N° casos grupos								
	1	2	3	4	5	6	8	9	17
25-34	0	1							
35-44	0	2	4	3					
45-54	0	2	4	6	12	10	12		
55-64			4	12	12	10	12	8	18
65-74	0	4	4	6	8	10	6		
75+	0	7	6						

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.

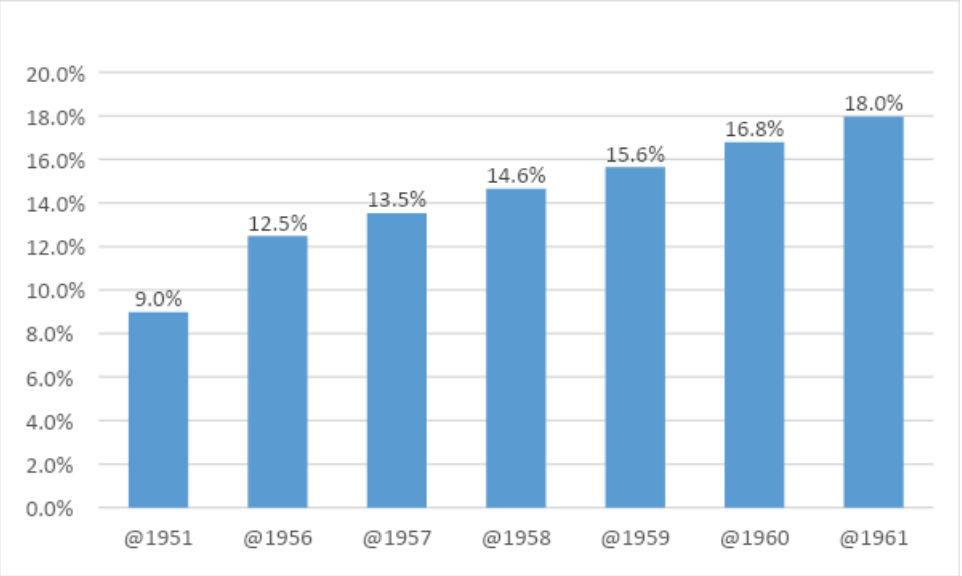
En este caso a las columnas tenemos la variable “continentes”.

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

Year	N.Amer	Europe	Asia	S.Amer	Oceania	Africa	Mid.Amer
1951	9,8%	9,0%	6,6%	9,4%	9,0%	0,9%	9,4%
1956	12,9%	12,5%	10,8%	13,2%	12,9%	13,6%	12,4%
1957	13,9%	13,5%	12,0%	13,9%	13,7%	14,9%	13,1%
1958	14,7%	14,6%	15,3%	14,7%	14,6%	16,0%	14,2%
1959	15,4%	15,6%	15,7%	15,5%	15,6%	17,0%	15,5%
1960	16,3%	16,8%	18,9%	16,2%	16,6%	18,3%	17,1%
1961	17,1%	18,0%	20,8%	17,2%	17,5%	19,3%	18,3%
TOTAL	**100,0%**	**100,0%**	**100,0%**	**100,0%**	**100,0%**	**100,0%**	**100,0%**

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?



Tan solo nos informa la tendencia creciente en número de móviles a lo largo de la serie analizada. 1961 acumula el mayor número de móviles activos en el continente europeo. Esta tendencia creciente, se constata en la tabla, que se da en todos los Continentes. No es una información muy valiosa y que de mucho juego para el estudio.

Frecuencias Condicionadas por filas (años)

Calcula las frecuencias condicionadas por filas.

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

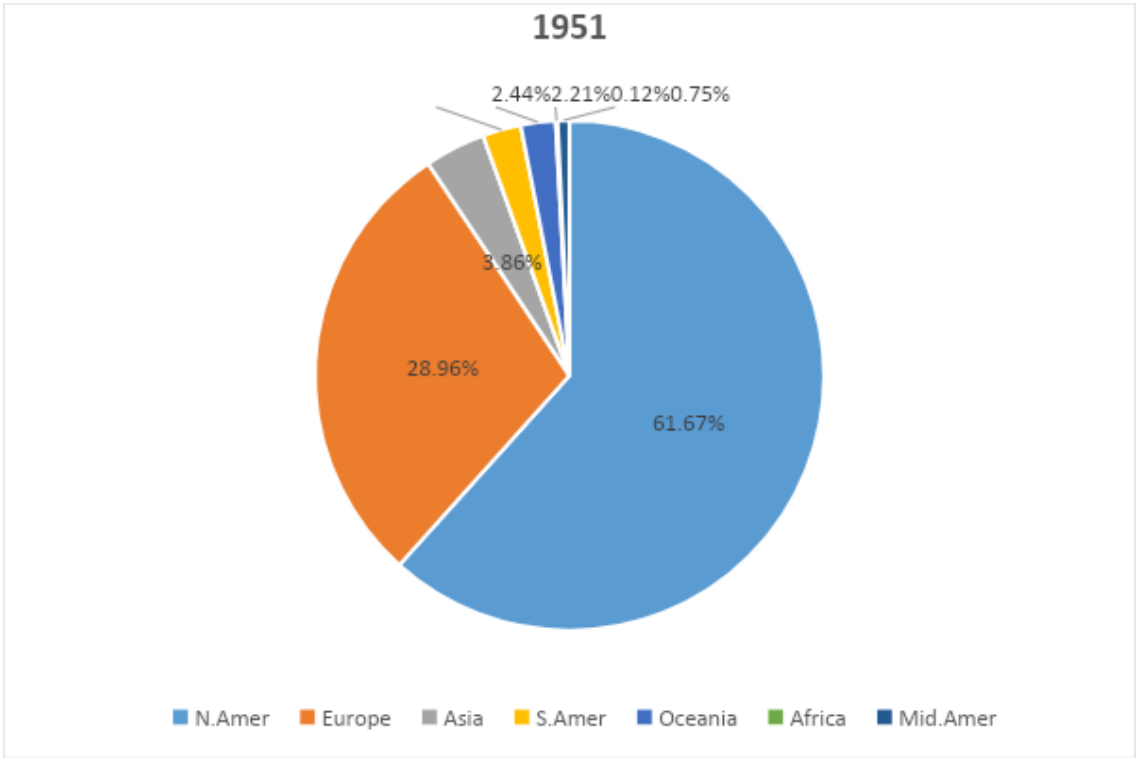
Para cada año calcularemos el porcentaje de cada continente. (Yo creo que tiene más sentido este cálculo en filas que en columnas, ¿verdad?)

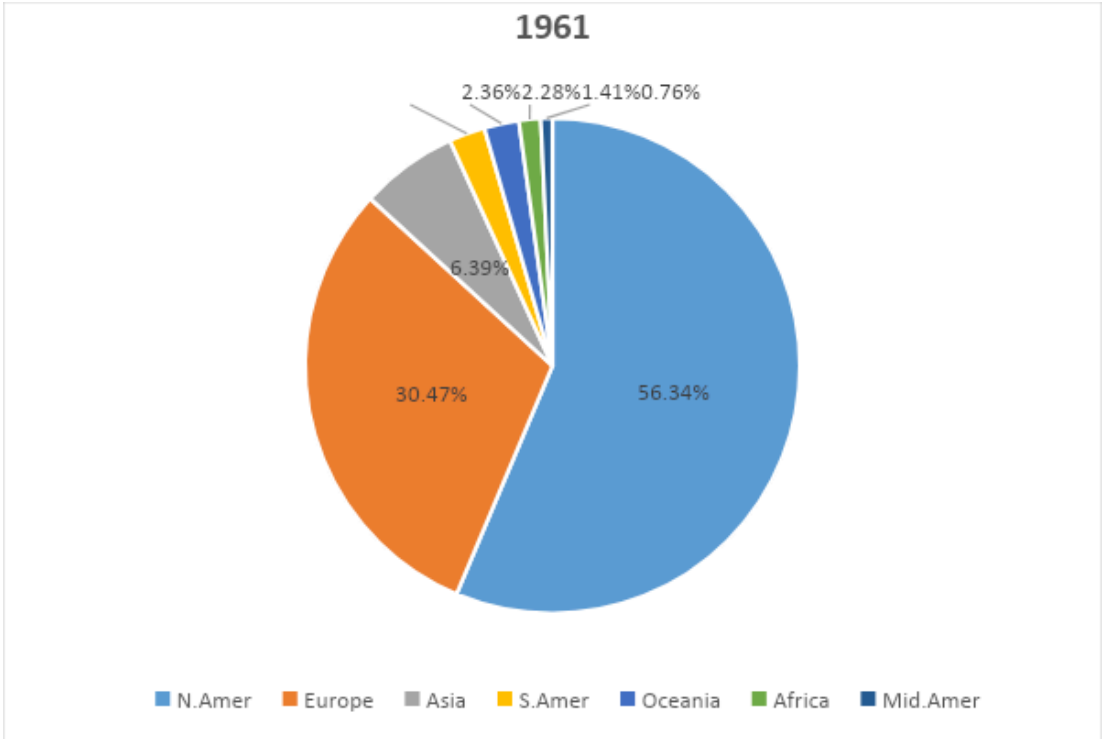
Year	N.Amer	Europe	Asia	S. Amer	Oceania	Africa	Mid. Amer	TOTAL
1951	61,67%	28,96%	3,86%	2,44%	2,21%	0,12%	0,75%	**100,0%**
1956	59,12%	29,34%	4,61%	2,51%	2,32%	1,38%	0,72%	**100,0%**
1957	58,84%	29,55%	4,75%	2,45%	2,30%	1,41%	0,70%	**100,0%**
1958	57,84%	29,75%	5,63%	2,40%	2,27%	1,40%	0,71%	**100,0%**
1959	57,53%	30,13%	5,49%	2,40%	2,30%	1,42%	0,73%	**100,0%**
1960	56,87%	30,17%	6,15%	2,35%	2,28%	1,42%	0,75%	**100,0%**
1961	56,34%	30,47%	6,39%	2,36%	2,28%	1,41%	0,76%	**100,0%**

Es sin duda más interesante. Nos permite constatar que la mayor concentración de móviles se da en América, da igual el año que observemos. Y que América Central mantiene una modesta presencia en este ranking por continentes en la variable que nos ocupa.

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.


1951
2.44%
2.21%
0.12%
0.75%
3.86%
28.96%
61.67%
N.Amer
Europe
Asia
S.Amer
Oceania
Africa
Mid.Amer


1961
2.36%
2.28%
1.41%
0.76%
6.39%
30.47%
56.34%
N.Amer
Europe
Asia
S.Amer
Oceania
Africa
Mid.Amer

Interesante ver cómo Europa y Asia van ganando terreno en la tenencia de móviles entre ambos años, restándole poco a poco el protagonismo inicial de Norteamérica.

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

¡Bien!

¡Vamos a seguir avanzando!