

Reporte estadístico de calidad de vida

Alan Padilla Campoy

2025-01-09

Contents

Introducción	1
Limpieza y validación	1
Exploración	2
Población por subgrupos diagnósticos	2
Análisis demográfico	3
Exploración de la variable ql por subgrupos diagnósticos	6
Exploración de la relación de alcohol, contraceptive_kind y education con ql	7
Análisis	10
Análisis de correlación	10
Análisis estadístico	10

Introducción

Este reporte analiza y explora datos sobre la calidad de vida de mujeres con enfermedades de la mama, evaluados mediante los cuestionarios de calidad de vida de la EORTC. En particular, se utilizan el EORTC-QLQ-C30, diseñado para cáncer en general, y el EORTC-QLQ-BR23, específico para cáncer de mama. El objetivo es identificar cómo factores demográficos, clínicos y de estilo de vida influyen en el bienestar de las pacientes.

En este reporte, se examina la relación del estado global de salud (*ql*) con estos factores. Los datos se extrajeron del conjunto de datos contenido en el archivo **data.xlsx**. El reporte se divide en dos secciones: una exploración de datos, que incluye estadísticas descriptivas y visualizaciones para ofrecer una visión general de la población, y un análisis estadístico que examina las relaciones entre estos factores y las métricas de calidad de vida.

Limpieza y validación

Se eliminaron las filas (entradas) con más del 20% de valores faltantes (NAs) y las columnas (variables) con más del 20% de NAs para asegurar la calidad y confiabilidad de los datos. Esto para evitar que la presencia excesiva de datos faltantes pudiera distorsionar los análisis posteriores. Las variables eliminadas fueron: *birth_number*, *lateral*, *histotype*, *gradeinv*, *erstatus*, *prstatus*, *her2status*, *brsee*, *brhl*.

Exploración

Población por subgrupos diagnósticos

Esta sección muestra de manera gráfica la cantidad de pacientes con cada diagnóstico.

En la base de datos tenemos 661 casos de Cáncer de mama , 60 casos de CDIS , 269 casos de Fibroadenoma y 477 casos de Otras enf. de la mama. Esto se observa en la siguiente gráfica.

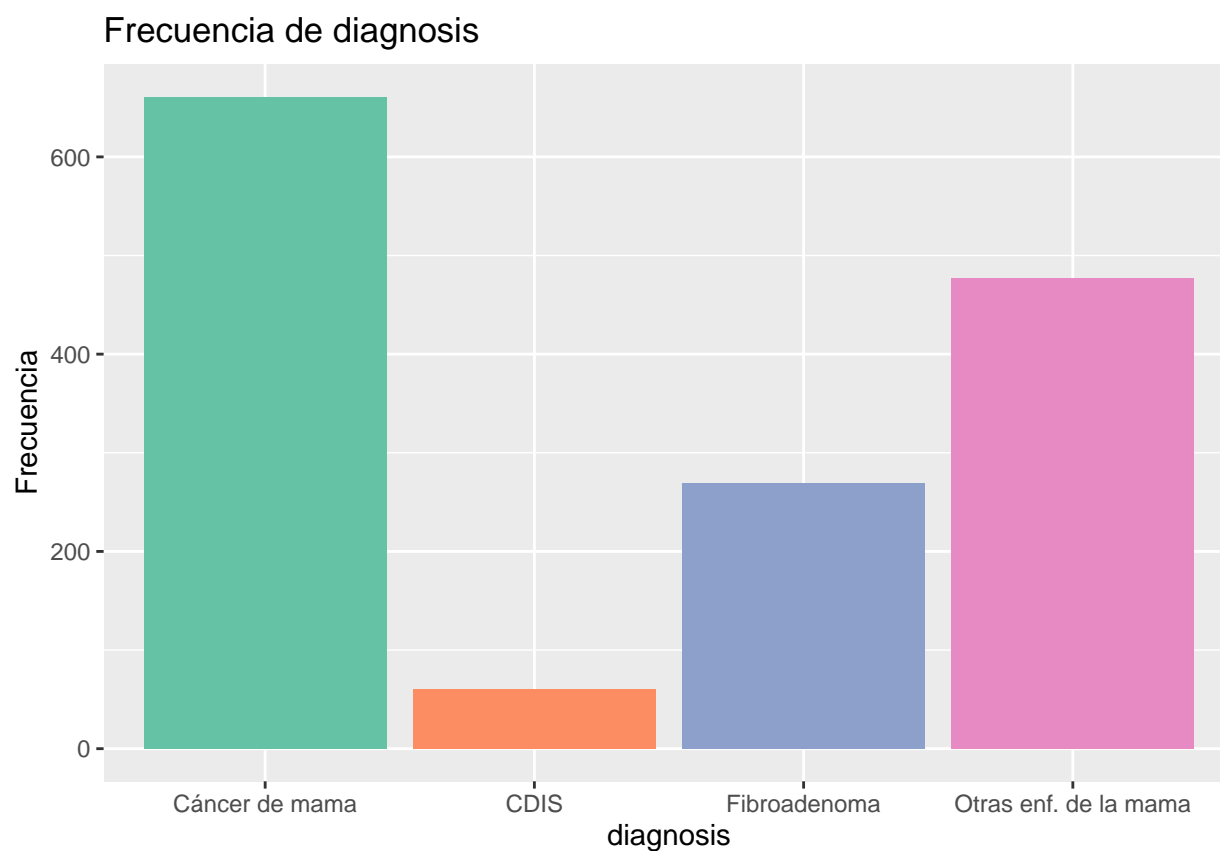


Figure 1: Distribución de pacientes según diagnóstico.

Análisis demográfico

Esta sección explora como estan distribuidos los datos sociodemográficos de las pacientes en la base de datos. Se analizan las distribuciones de edad (*age*), estado civil (*marital_status*), nivel educativo (*education*), consumo de alcohol (*alcohol*) y tabaquismo (*smokingstatus*).

Edad al momento del registro

Las **edades** de las participantes oscilan entre 15 y 87, con un primer cuartil de 36, una mediana de 46, una media de 47.01 y un tercer cuartil de 57. La distribución se puede apreciar en la siguiente figura.

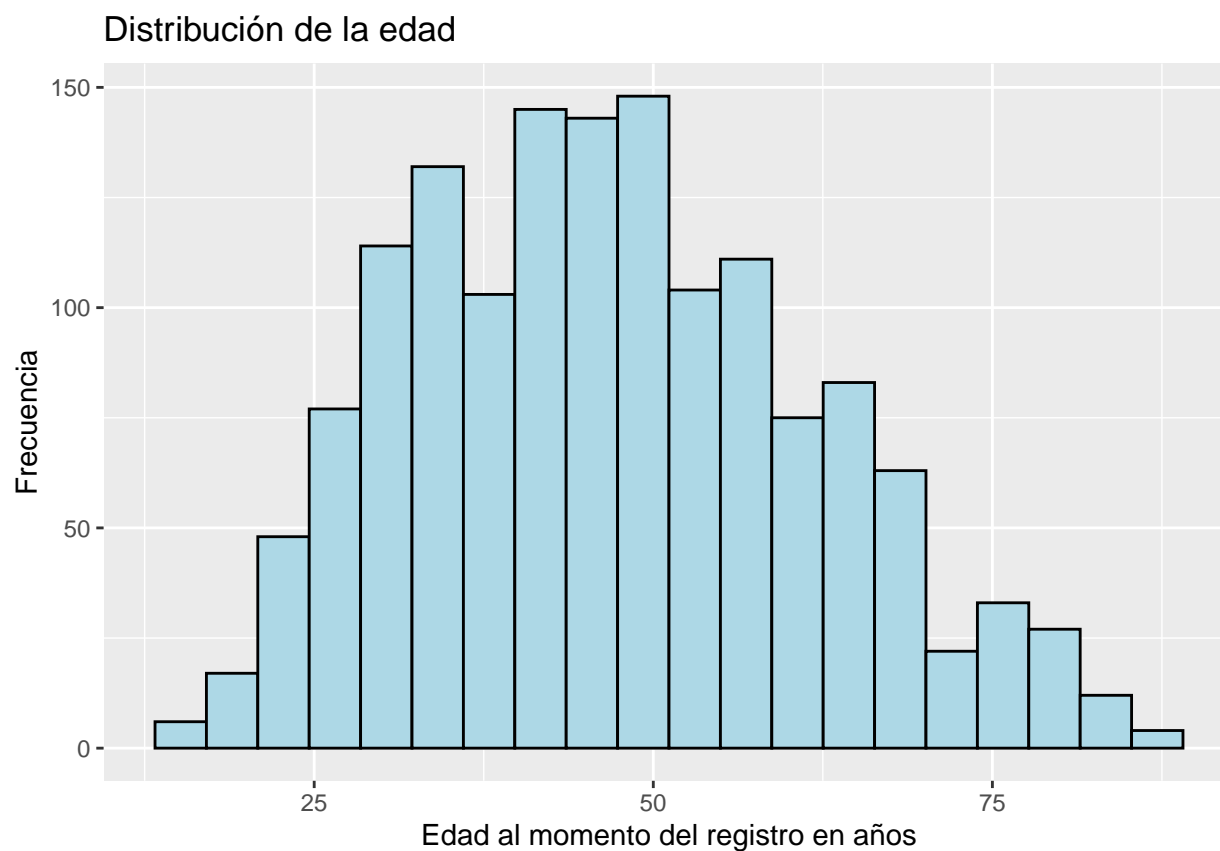


Figure 2: Distribución de las edades de las pacientes.

Estado civil de las pacientes

Los **estados civiles** de las participantes se distribuyen de la siguiente manera: 281 “Soltera”, 967 “Casada/En una relación”, 127 “Divorciada/Separada”, y 69 “Viuda”.

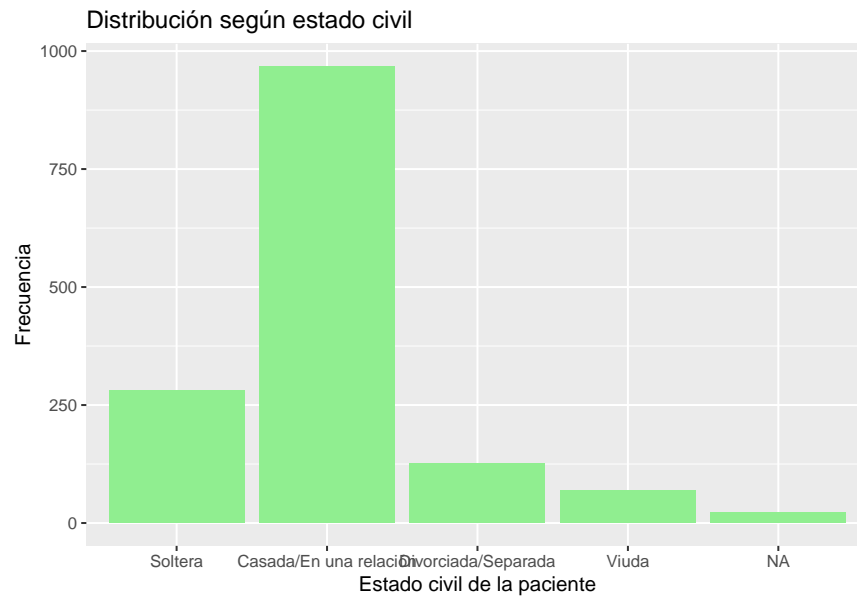


Figure 3: Distribución de los estados civiles de las pacientes.

Nivel educativo de las pacientes

Los **niveles educativos** se distribuyen de la siguiente manera: 134 con educación “Bajo”, 346 con educación “Medio”, y 983 con educación “Alto”.

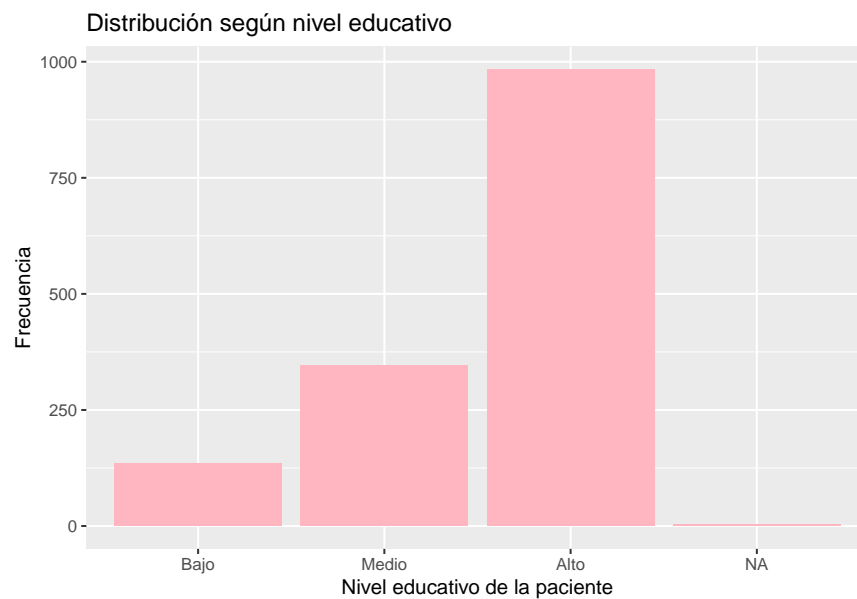


Figure 4: Distribución de los niveles educativos de las pacientes.

Consumo de alcohol de las pacientes

La frecuencia de **consumo de alcohol** muestra que 432 no consumen alcohol (“Ninguno”), 810 lo hacen “Ocasionalmente”, 184 consumen “Semanalmente”, y 41 lo hacen “Diariamente”.

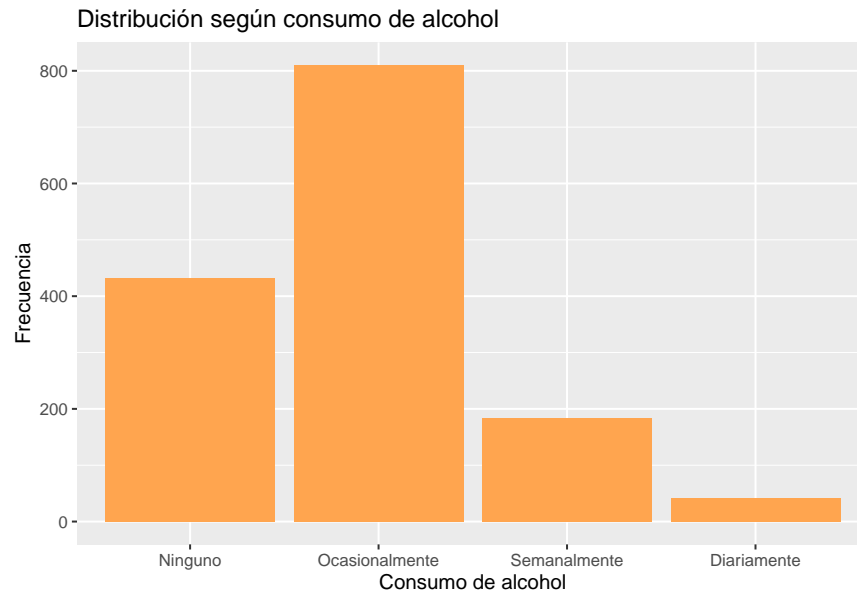


Figure 5: Distribución según el consumo de alcohol de las pacientes.

Tabaquismo

En cuanto al **tabaquismo**, 995 nunca han fumado (“No”), 297 actualmente fuman (“Sí”), y 175 son exfumadoras (“Exfumador”).

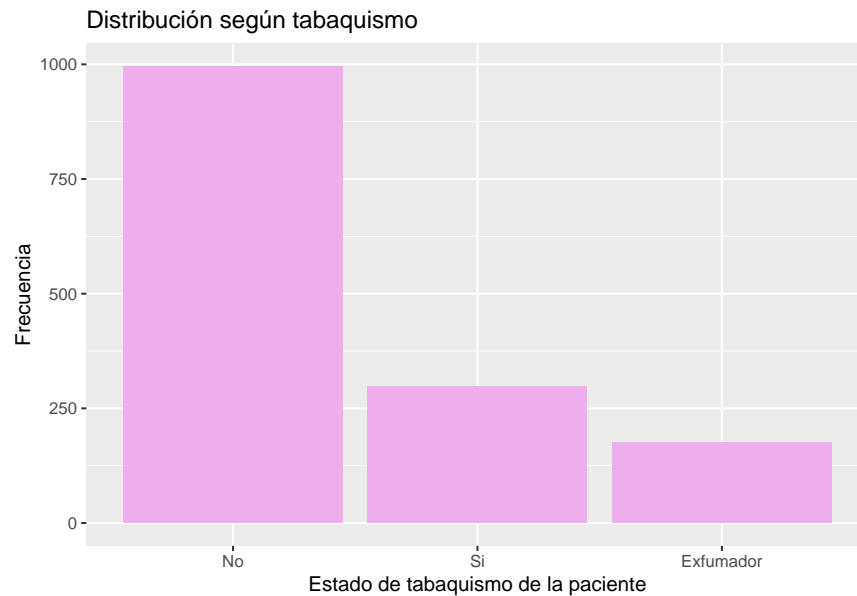


Figure 6: Distribución del estado de tabaquismo de las pacientes.

Exploración de la variable ql por subgrupos diagnósticos

La siguiente tabla muestra el número de pacientes (n), las medias, medianas y la desviación estándar del estado global de salud (ql) según el tipo de enfermedad.

Table 1: Estadísticas de ql por diagnóstico

diagnosis	n	media	mediana	desviacion
Cáncer de mama	661	62.38023	66.66667	23.13114
CDIS	60	64.86111	66.66667	24.74406
Fibroadenoma	269	72.61462	75.00000	21.21515
Otras enf. de la mama	477	71.76799	75.00000	20.10727

El siguiente gráfico muestra un gráfico de cajas del estado global de salud (ql) según el diagnóstico.

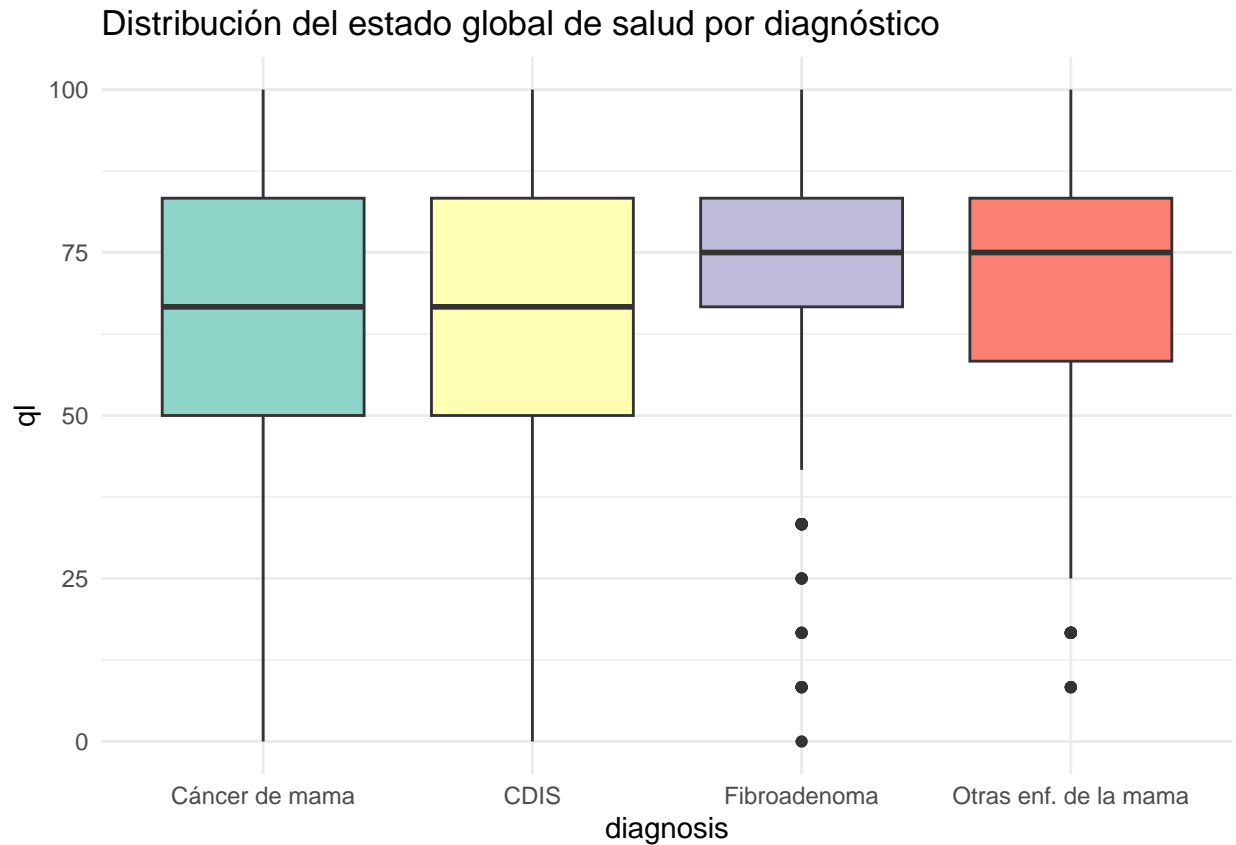


Figure 7: Distribución del estado global de salud (ql) por diagnóstico

Exploración de la relación de alcohol, contraceptive_kind y education con ql

En esta sección exploramos el cambio del estado global de salud (ql) en subgrupos creados a partir de tres de las variables independientes categoricas. Estas variables son: frecuencia de consumo de alcohol ($alcohol$), método anticonceptivo utilizado ($contraceptive_kind$) y nivel educativo del paciente ($education$).

frecuencia de consumo de alcohol ($alcohol$)

La siguiente tabla muestra el numero de pacientes (n), las medias, medianas y la desviación estándar del estado global de salud (ql) según frecuencia de consumo de alcohol ($alcohol$).

Table 2: Estadísticas de ql por alcohol

alcohol	n	media	mediana	desviacion
Ninguno	432	62.88580	66.66667	24.37368
Ocasionalmente	810	68.45679	66.66667	21.35899
Semanalmente	184	73.50543	75.00000	20.57058
Diariamente	41	67.07317	66.66667	20.40827

El siguiente gráfico muestra un gráfico de cajas del estado global de salud (ql) según frecuencia de consumo de alcohol ($alcohol$).

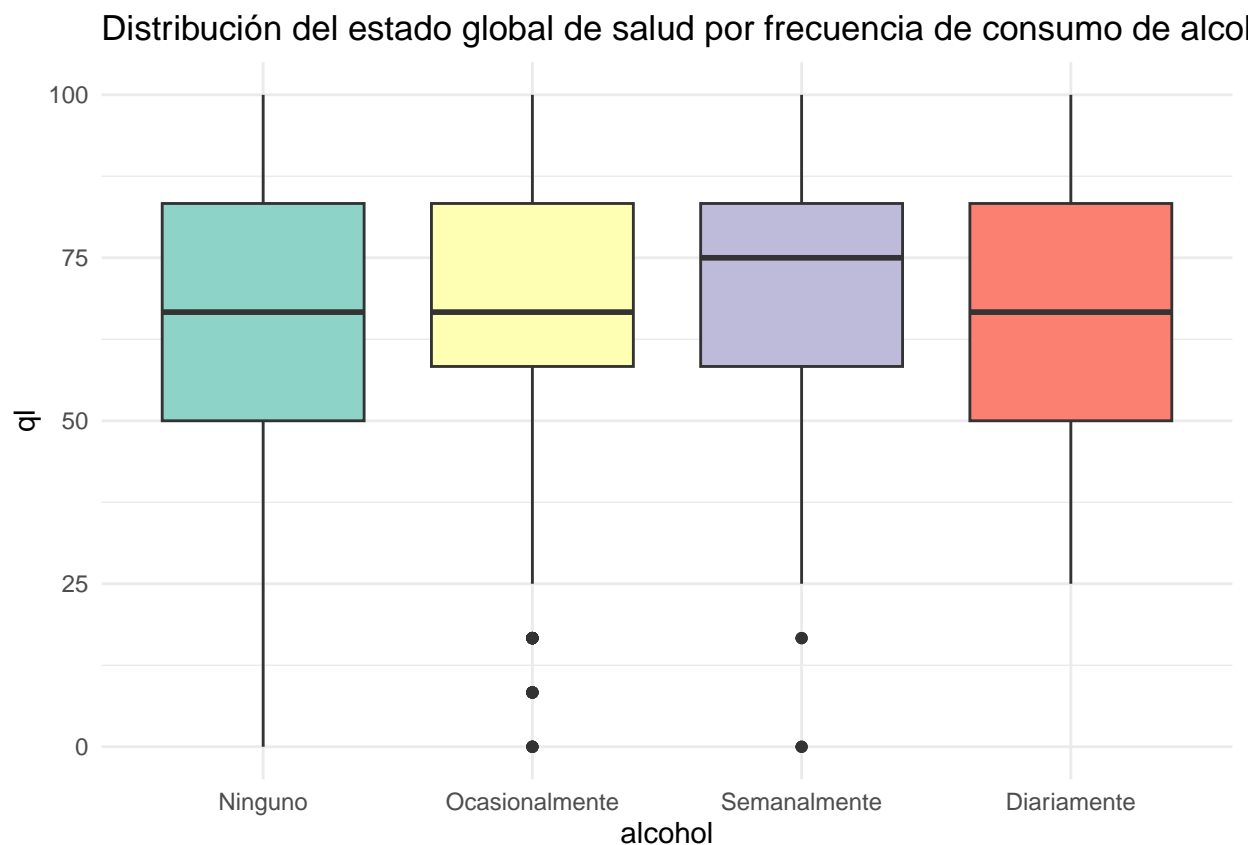


Figure 8: Distribución del estado global de salud (ql) por frecuencia de consumo de alcohol ($alcohol$)

método anticonceptivo utilizado (*contraceptive_kind*)

La siguiente tabla muestra el numero de pacientes (n), las medias, medianas y la desviación estándar del estado global de salud (*ql*) según método anticonceptivo utilizado (*contraceptive_kind*).

Table 3: Estadísticas de *ql* por *contraceptive_kind*

contraceptive_kind	n	media	mediana	desviacion
Ninguno	1163	66.15076	66.66667	22.53525
AC oral	126	71.95767	75.00000	21.28127
Inyección	1	83.33333	83.33333	NA
Anillo vaginal	15	76.66667	75.00000	13.43710
Parche	1	41.66667	41.66667	NA
DIU hormonal	62	72.58065	75.00000	21.14586
DIU cobre	31	75.53763	75.00000	18.74975
Otro	68	70.22059	75.00000	23.75927

El siguiente gráfico muestra un gráfico de cajas del estado global de salud (*ql*) según método anticonceptivo utilizado (*contraceptive_kind*).

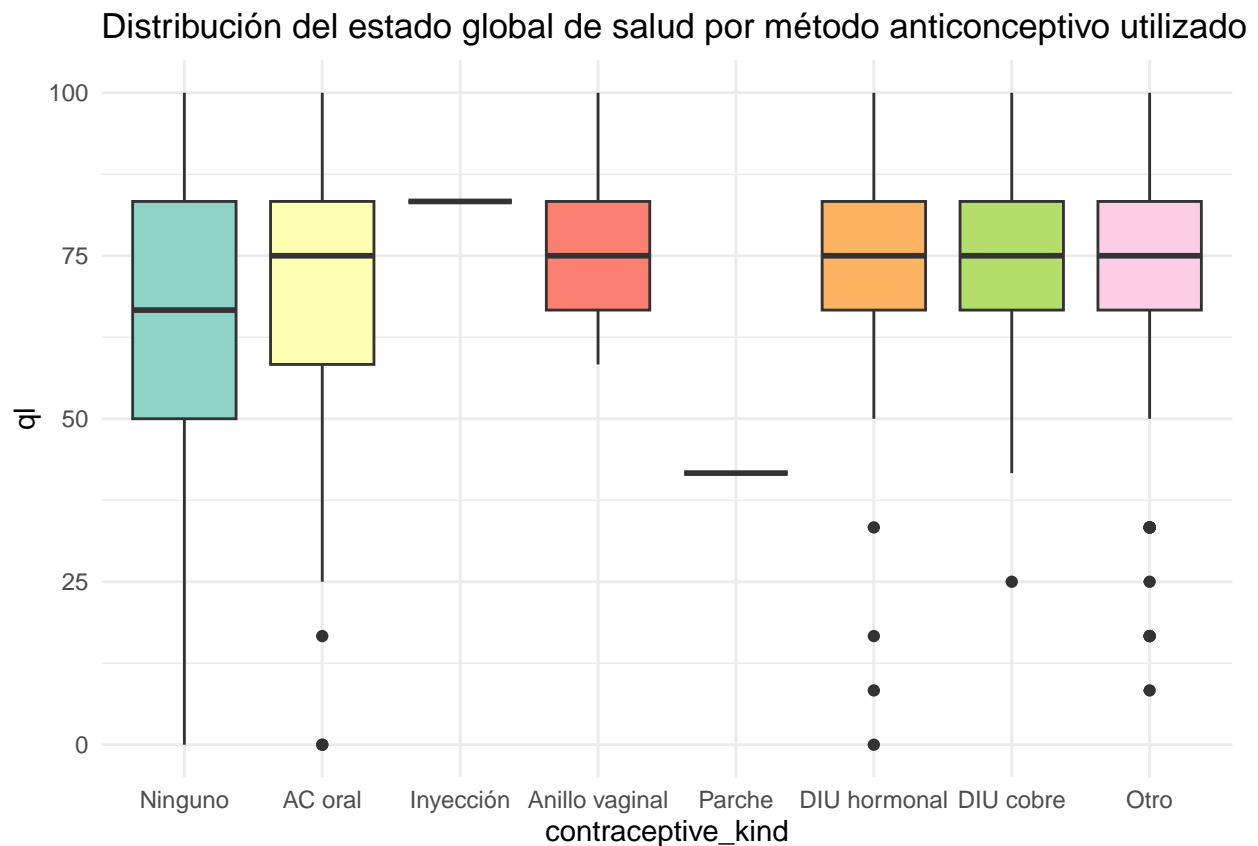


Figure 9: Distribución del estado global de salud (*ql*) por método anticonceptivo utilizado (*contraceptive_kind*)

nivel educativo del paciente (*education*)

La siguiente tabla muestra el numero de pacientes (*n*), las medias, medianas y la desviación estándar del estado global de salud (*ql*) según nivel educativo del paciente (*education*).

Table 4: Estadísticas de ql por education

education	n	media	mediana	desviacion
Bajo	134	57.46269	50.00000	24.27453
Medio	346	62.98170	66.66667	24.05770
Alto	983	70.33740	75.00000	20.75642
NA	4	64.58333	79.16667	45.32547

El siguiente gráfico muestra un gráfico de cajas del estado global de salud (*ql*) según nivel educativo del paciente (*education*).

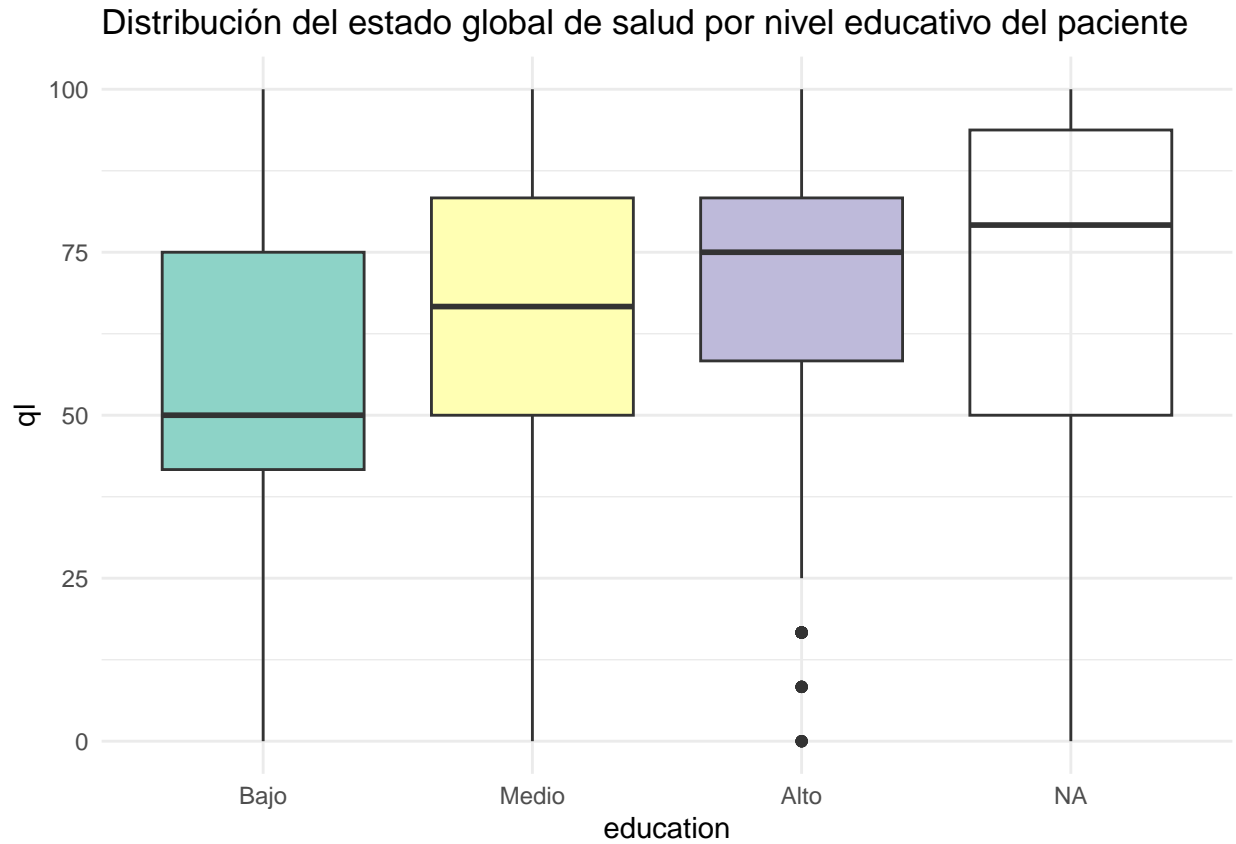


Figure 10: Distribución del estado global de salud (*ql*) por nivel educativo del paciente (*education*)

Análisis

Análisis de correlación

En esta sección se realizó un análisis de las correlaciones del estado global de salud (*ql*) y las variables independientes numéricas (*age*, *weight*, *height*, *bmi*). Las correlaciones son las siguientes:

Table 5: Correlaciones entre *ql* y las variables numéricas

	Variable	Correlacion
<i>ql</i>	<i>ql</i>	1.0000000
<i>age</i>	<i>age</i>	-0.2015391
<i>bmi</i>	<i>bmi</i>	-0.1859559
<i>weight</i>	<i>weight</i>	-0.1435425
<i>height</i>	<i>height</i>	0.1045577

Considerando que las correlaciones fuertes son mayores a 0.7 y las débiles a 0.3 podemos encontrar la fuerza de la relación entre *ql* y las variables independientes numéricas. En este caso todas las correlaciones son débiles.

Análisis estadístico

Variables significativas

Después de realizar diversas pruebas estadísticas para comparar diferentes grupos, hay evidencia significativa para inferir que las variables **marital_status** (estado civil de la paciente), **education** (nivel educativo del paciente), **alcohol** (frecuencia de consumo de alcohol), **smokingstatus** (estado de tabaquismo del paciente), **bust** (medida del busto en cm), **menopause_yn** (estado de menopausia del paciente), **pregnancy_number** (número de embarazos), **comorb_none** (el paciente no presenta comorbilidades), **comorb_heart** (comorbilidad: enfermedades cardíacas), **comorb_hypertension** (comorbilidad: hipertensión arterial), **comorb_paod** (comorbilidad: enfermedad arterial periférica), **comorb_lung** (comorbilidad: enfermedades pulmonares), **comorb_diabetes** (comorbilidad: diabetes), **comorb_kidneys** (comorbilidad: enfermedad renal), **comorb_liver** (comorbilidad: enfermedades hepáticas), **comorb_stroke** (comorbilidad: secuelas de un ACV), **comorb_neurological** (comorbilidad: enfermedades neurológicas), **comorb_cancerlast5years** (cáncer en los últimos cinco años), **comorb_depression** (comorbilidad: depresión), **comorb_gastrointestinal** (comorbilidad: enfermedades gastrointestinales), **comorb_endometriosis** (comorbilidad: endometriosis), **comorb_arthritis** (comorbilidad: artritis), **comorb_incontinence** (comorbilidad: incontinencia urinaria), **comorb_uti** (comorbilidad: infecciones del tracto urinario), **cancer_breast** (cáncer de mama previo), **contraceptive_kind** (método anticonceptivo utilizado), **pre_op** (cirugía previa en las mamas), **cancer_kind_family_1** (historia familiar de cáncer de mama o de ovario), **breastcancer_first** (primera enfermedad mamaria del paciente), **diagnosis** (diagnóstico) pueden afectar la métrica del estado global de salud (*ql*).

Efecto por niveles

Variable: **marital_status**

La diferencia entre Divorciada/Separada y Casada/En una relación es significativa ($p = 0.0054$).

La diferencia entre Viuda y Soltera es significativa ($p = 0.0071$).

La diferencia entre Viuda y Casada/En una relación es significativa ($p = 0.0011$).

Variable: **education**

La diferencia entre Medio y Bajo es significativa ($p = 0.0323$).

La diferencia entre Alto y Bajo es significativa ($p = 0$).

La diferencia entre Alto y Medio es significativa ($p = 0$).

Variable: alcohol

La diferencia entre Ocasionalmente y Ninguno es significativa ($p = 5e-04$).

La diferencia entre Semanalmente y Ninguno es significativa ($p = 0$).

La diferencia entre Semanalmente y Ocasionalmente es significativa ($p = 0.0174$).

Variable: smokingstatus

La diferencia entre Si y No es significativa ($p = 0.0116$).

Variable: bust

La diferencia entre 95 y 75 es significativa ($p = 0.0171$).

Variable: menopause_yn

La diferencia entre Si y No es significativa ($p = 0$).

Variable: pregnancy_number

La diferencia entre 2 y 0 es significativa ($p = 0.0209$).

La diferencia entre 5 y 0 es significativa ($p = 0.0114$).

Variable: comorb_none

La diferencia entre Si y No es significativa ($p = 0$).

Variable: comorb_heart

La diferencia entre Si y No es significativa ($p = 2e-04$).

Variable: comorb_hypertension

La diferencia entre Si y No es significativa ($p = 0$).

Variable: comorb_paod

La diferencia entre Si y No es significativa ($p = 0$).

Variable: comorb_lung

La diferencia entre Si y No es significativa ($p = 0$).

Variable: comorb_diabetes

La diferencia entre Si y No es significativa ($p = 1e-04$).

Variable: comorb_kidneys

La diferencia entre Si y No es significativa ($p = 1e-04$).

Variable: comorb_liver

La diferencia entre Si y No es significativa ($p = 0.0081$).

Variable: comorb_stroke

La diferencia entre Si y No es significativa ($p = 0.009$).

Variable: comorb_neurological

La diferencia entre Si y No es significativa ($p = 0$).

Variable: comorb_cancerlast5years

La diferencia entre Si y No es significativa ($p = 0$).

Variable: comorb_depression

La diferencia entre Si y No es significativa ($p = 0$).

Variable: comorb_gastrointestinal

La diferencia entre Si y No es significativa ($p = 0$).

Variable: comorb_endometriosis

La diferencia entre Si y No es significativa ($p = 0.0063$).

Variable: comorb_arthritis

La diferencia entre Si y No es significativa ($p = 0$).

Variable: comorb_incontinence

La diferencia entre Si y No es significativa ($p = 0$).

Variable: comorb_uti

La diferencia entre Si y No es significativa ($p = 0$).

Variable: cancer_breast

La diferencia entre Si y No es significativa ($p = 0$).

Variable: pre_op

La diferencia entre Si y No es significativa ($p = 0$).

Variable: cancer_kind_family_1

La diferencia entre Si y No es significativa ($p = 0.0171$).

Variable: breastcancer_first

La diferencia entre Si y No es significativa ($p = 0.0203$).

Variable: diagnosis

La diferencia entre Fibroadenoma y Cáncer de mama es significativa ($p = 0$).

La diferencia entre Otras enf. de la mama y Cáncer de mama es significativa ($p = 0$).