

ANÀLISI DE LA BASE DE DADES PREDIMED



Universitat de Lleida

Bioestadística

Doble grau en fisioteràpia, nutrició humana i
dietètica

Curs 2019-2020

Meritxell Creus Graells i Núria Sitjà Gras

Taula de contingut

Anàlisi per regressió logística:	4
Anàlisi de supervivència:	9
Conclusions:.....	11

Taula d'il·lustracions

Il·lustració 1:CrossTable de group i event	4
Il·lustració 2:Taula representa el grup en funció de patir l'event i les diferents variables que interaccionen.....	5
Il·lustració 3: Anova dels diversos models.....	6
Il·lustració 4: Resultat regressió logística base dd.....	6
Il·lustració 5: Resultat regressió logística base d	6
Il·lustració 6: Representació dels models	7
Il·lustració 7: Odds ratio i IC model 2	8
Il·lustració 8: Diferència entre dones i homes de patir l'event amb les mateixes variables.....	8
Il·lustració 9: Anàlisi de cox en funció de l'aparició de l'event i el grup	9
Il·lustració 10: Significació	9
Il·lustració 11: Representació gràfica de les diferents variables nominals	10
Il·lustració 14: Hazard ratio de les variables significatives	11
Il·lustració 13: Anàlisi de cox	11

Introducció:

L'estudi PREDIMED, es un estudi que es va realitzar entre el 2003 i el 2010. Es va estudiar la relació entre l'aparició de malalties cardiovasculars envers el tipus de dieta que se'ls hi va assignar als participants.

Es volia estudiar la relació entre patir malalties cardiovasculars amb un tipus de dieta concreta que els hi van assignar. En aquest estudi van participar 7.000 pacients amb un alt risc cardiovascular, que se'ls va distribuir en tres grups; grup d'AOVE amb dieta mediterrània, grup de nous amb dieta mediterrània i un grup control. A aquests participants se'ls hi feia un seguiment periòdicament cada 6 mesos, a més de sessions educatives segons el grup que pertanyien. Destacar que els grups que formaven part del grup experimental se'ls hi va subministrar gratuïtament o nous.

Durant aquest temps es van mesurar també les següents variables; el sexe, l'edat, si eren fumadors, si ho havien deixat o si no havien fumat mai, el BMI, la circumferència de la cintura, la relació entre cintura i altura, si tenien hipertensió, si eren diabètics, els nivells de colesterol, l'historial familiar, els nivells hormonals, l'adherència a la dieta, el temps a presentar la malaltia cardiovascular i finalment si aquesta apareixia o no.

Finalment es va concloure que la dieta mediterrània complementada amb AVOE o nous, tenia una menor taxa d'esdeveniments cardiovasculars. En aquest treball pretenem fer un anàlisi estadístic per corroborar que aquesta afirmació continguda en el treball es correcta. A més determinar quines variables influeixen més en l'aparició de malalties cardiovasculars mitjançant regressió logística i finalment es realitzarà un anàlisi de supervivència mitjançant la tècnica de cox.

Anàlisi per regressió logística:

L'estudi PREDIMED és un estudi de cohorts observacional, es per aquest fet que a continuació realitzem un anàlisi per regressió logística.

Cell Contents

	Count	Row Percent	Column Percent	Adj Std Resid
--	-------	-------------	----------------	---------------

Total Observations in Table: 5661

event	group			Row Total
	Control	MedDiet + Nuts	MedDiet + VOO	
No	1752	1806	1879	5437
	32.224%	33.217%	34.559%	96.043%
	95.114%	96.836%	96.162%	
	-2.490	2.146	0.332	
Yes	90	59	75	224
	40.179%	26.339%	33.482%	3.957%
	4.886%	3.164%	3.838%	
	2.490	-2.146	-0.332	
Column Total	1842	1865	1954	5661
	32.538%	32.945%	34.517%	

Statistics for All Table Factors

Pearson's Chi-squared test

Chi² = 7.345175 d.f. = 2 p = 0.02541063

Minimum expected frequency: 72.88606

Il·lustració 1: CrossTable de group i event

La il·lustració 1 és un CrossTable que relaciona les dues variables principals, essent el risc de patir complicacions cardiovasculars (event) i els diferents grups de dietes (group). El valor $p=0.02$ és significatiu i indica els percentatges observats no són compatibles amb la hipòtesi d'igualtat entre els diferents grups en funció de l'aparició de malalties cardiovasculars.

El valor ajustat del grup control en relació amb no patir complicacions cardiovascular és -2,490, i estableix que la probabilitat de no patir les complicacions és més baixa en el grup control, i que en canvi, el fet de patir complicacions cardiovasculars augmenta en el grup control, ja que el valor ajustat és de 2.490.

D'altra banda, en el grup de la dieta mediterrània suplementada amb nous ocorre lo contrari, el valor ajustat en relació amb no patir complicacions cardiovasculars és de 2.146 i significa que augmenta en aquest grup; en canvi, la probabilitat de no patir aquestes complicacions en aquest grup disminueix, ja que el valor ajustat es de -2.146, podent ser un factor protector.

Tot i que, pel que fa al grup de la dieta mediterrània suplementada amb AOVE, no es pot establir que de forma significativa influeixi sobre la hipòtesi, perquè el valor ajustat és menor de 1.96.

Així doncs, amb el CrossTable es pot establir que la dieta mediterrània amb suplementació de nous suporti la hipòtesi que la dieta és un factor protector.

L'estudi PREDIMED inclou més variables: sex, age, smoking, BMI, waist circumference, waist-to-height ratio, hypertension, type-2 diabetes, dyslipidemia, hormone replacement therapy, MedDiet Adherence score, follow-up to main event. La il·lustració 2 inclou totes les variables presents en l'anàlisi de l'estudi més la variable edat (és la age desglossada en diferents franges d'edat), amb la finalitat d'observar la relació entre les variables estudiades i l'aparició de complicacions cardiovascular en funció de la dieta predeterminada, per poder establir quines són un factor de risc per a desenvolupar-les.

	No				Yes			
	Control N=1945	MedDiet + Nuts N=2030	MedDiet + VOO N=2097	p. overall	Control N=97	MedDiet + Nuts N=70	MedDiet + VOO N=85	p. overall
Sex:				<0.001				0.947
Male	754 (38.8%)	927 (45.7%)	847 (40.4%)		58 (59.8%)	41 (58.6%)	52 (61.2%)	
Female	1191 (61.2%)	1103 (54.3%)	1250 (59.6%)		39 (40.2%)	29 (41.4%)	33 (38.8%)	
Age	67.2 (6.24)	66.6 (5.98)	66.9 (6.17)	0.003	69.2 (6.84)	69.6 (6.34)	69.3 (6.76)	0.926
Smoking:				0.236				0.361
Never	1238 (63.7%)	1224 (60.3%)	1316 (62.8%)		44 (45.4%)	35 (50.0%)	35 (41.2%)	
Current	256 (13.2%)	281 (13.8%)	272 (13.0%)		14 (14.4%)	15 (21.4%)	20 (23.5%)	
Former	451 (23.2%)	525 (25.9%)	509 (24.3%)		39 (40.2%)	20 (28.6%)	30 (35.3%)	
Body mass index	30.3 (3.97)	29.7 (3.75)	30.0 (3.71)	<0.001	30.3 (3.88)	29.2 (4.33)	29.6 (3.58)	0.168
waist circumference	101 (10.7)	100 (10.6)	100 (10.5)	0.079	103 (11.5)	101 (10.1)	102 (10.0)	0.341
waist-to-height ratio	0.63 (0.07)	0.62 (0.06)	0.63 (0.07)	<0.001	0.64 (0.07)	0.63 (0.07)	0.63 (0.06)	0.603
Hypertension:				0.484				0.042
No	322 (16.6%)	348 (17.1%)	377 (18.0%)		9 (9.28%)	14 (20.0%)	19 (22.4%)	
Yes	1623 (83.4%)	1682 (82.9%)	1720 (82.0%)		88 (90.7%)	56 (80.0%)	66 (77.6%)	
Type-2 diabetes:				0.028				0.591
No	1035 (53.2%)	1123 (55.3%)	1073 (51.2%)		37 (38.1%)	27 (38.6%)	27 (31.8%)	
Yes	910 (46.8%)	907 (44.7%)	1024 (48.8%)		60 (61.9%)	43 (61.4%)	58 (68.2%)	
Dyslipidemia:				0.573				0.245
No	529 (27.2%)	534 (26.3%)	582 (27.8%)		34 (35.1%)	27 (38.6%)	40 (47.1%)	
Yes	1416 (72.8%)	1496 (73.7%)	1515 (72.2%)		63 (64.9%)	43 (61.4%)	45 (52.9%)	
Family history of premature CHD:				0.698				0.480
No	1503 (77.3%)	1581 (77.9%)	1610 (76.8%)		77 (79.4%)	59 (84.3%)	65 (76.5%)	
Yes	442 (22.7%)	449 (22.1%)	487 (23.2%)		20 (20.6%)	11 (15.7%)	20 (23.5%)	
Hormone-replacement therapy:				0.775				0.263
No	1721 (98.2%)	1777 (98.4%)	1843 (98.1%)		90 (100%)	58 (98.3%)	75 (100%)	
Yes	31 (1.77%)	29 (1.61%)	36 (1.92%)		0 (0.00%)	1 (1.69%)	0 (0.00%)	
MedDiet Adherence score	8.45 (1.94)	8.83 (1.90)	8.79 (1.97)	<0.001	8.18 (2.03)	8.26 (1.95)	8.31 (1.83)	0.900
Follow-up to main event (years)	4.17 (1.70)	4.35 (1.69)	4.70 (1.56)	<0.001	2.53 (1.66)	3.24 (1.69)	3.23 (1.83)	0.008
edat:				0.009				0.773
49-62anys	515 (26.5%)	579 (28.5%)	580 (27.7%)		18 (18.6%)	12 (17.1%)	18 (21.2%)	
62-67anys	487 (25.1%)	589 (29.0%)	581 (27.7%)		20 (20.6%)	15 (21.4%)	16 (18.8%)	
67-72anys	497 (25.6%)	480 (23.6%)	499 (23.8%)		18 (18.6%)	20 (28.6%)	20 (23.5%)	
72-87anys	445 (22.9%)	382 (18.8%)	436 (20.8%)		41 (42.3%)	23 (32.9%)	31 (36.5%)	

Il·lustració 2: Taula representa el grup en funció de patir l'event i les diferents variables que interaccionen

L'il·lustració 2 mostra com en el grup de persones que no presenten complicacions cardiovasculars en funció de la dieta, les variables sex, age, BMI, waist-to-height ratio, MedDiet Adherence score, follow-up to main event i l'edat, són significatives. En canvi, en el grup de persones que han presentat les complicacions cardiovasculars en funció de les diferents dietes, l'única variable que es mostra significativa és el follow-up to main event.

Analysis of Deviance Table

```

Model 1: event ~ age + diab + hyperchol + group
Model 2: event ~ sex + group + age + smoke + waist + diab + hyperchol +
p14
Model 3: event ~ group + sex + age + smoke + waist + wth + diab + hyperchol +
p14
Model 4: event ~ group + sex + age + smoke + waist + wth + htn + diab +
hyperchol + p14 + edat
Model 5: event ~ (group + sex + age + smoke + bmi + waist + wth + htn +
diab + hyperchol + famhist + hormo + p14 + toevent + edat) -
toevent
Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      5654      1813.2
2      5649      1764.4  5    48.807 2.43e-09 ***
3      5648      1759.6  1     4.766 0.02902 *
4      5644      1752.4  4     7.242 0.12364
5      5641      1748.8  3     3.515 0.31881
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Il·lustració 3: Anova dels diversos models

Mitjançant l'anàlisi d'Anova s'han buscat les millors variables predictores de l'aparició de complicacions cardiovasculars. S'han creat 5 models en funció de diferents variables predictores. La il·lustració 3 mostra que el millor model per estimar la probabilitat de què a una persona presenti complicacions cardiovasculars es el Model 2, així doncs són unes bones variables predictores sex, age, group, smoke, waist, diab, hyperchol i p14, ja que dona una p significativa i mostra menys residuals. Més endavant del treball s'explica perquè creiem que el millor model es el 2 i no el 3 tot i sers significatiu.

Regressió logística amb la base dd	Regressió logística amb la base d
<p>Call: glm(formula = event ~ sex + group + age + smoke + waist + diab + hyperchol + p14, family = "binomial", data = dd)</p> <p>Deviance Residuals: Min IQ Median 3Q Max -0.7924 -0.3115 -0.2388 -0.1838 3.0438</p> <p>Coefficients: Estimate Std. Error z value Pr(> z) (Intercept) -7.366427 1.164764 -6.324 2.54e-10 *** sexFemale -0.530847 0.186018 -2.854 0.00432 ** groupMedDiet + Nuts -0.421408 0.173955 -2.423 0.01541 * groupMedDiet + V00 -0.247758 0.162435 -1.525 0.12719 age 0.073408 0.011421 6.427 1.30e-10 *** smokeCurrent 0.509172 0.228524 2.228 0.02587 * smokeFormer 0.408400 0.195296 2.091 0.03651 * waist 0.003483 0.007008 0.497 0.61919 diabYes 0.483255 0.148544 3.253 0.00114 ** hypercholYes -0.323302 0.146775 -2.203 0.02762 * p14 -0.110236 0.035237 -3.128 0.00176 ** --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1</p> <p>(Dispersion parameter for binomial family taken to be 1)</p> <p>Null deviance: 1885.8 on 5659 degrees of freedom Residual deviance: 1764.4 on 5649 degrees of freedom AIC: 1786.4</p> <p>Number of Fisher Scoring iterations: 6</p> <p><i>Il·lustració 4: Resultat regressió logística base dd</i></p>	<p>Call: glm(formula = event ~ sex + group + age + smoke + waist + diab + hyperchol + p14, family = "binomial", data = d)</p> <p>Deviance Residuals: Min IQ Median 3Q Max -0.7925 -0.3115 -0.2388 -0.1837 3.0438</p> <p>Coefficients: Estimate Std. Error z value Pr(> z) (Intercept) -7.367026 1.164739 -6.325 2.53e-10 *** sexFemale -0.530874 0.186019 -2.854 0.00432 ** groupMedDiet + Nuts -0.421358 0.173954 -2.422 0.01543 * groupMedDiet + V00 -0.247713 0.162435 -1.525 0.12726 age 0.073418 0.011420 6.429 1.29e-10 *** smokeCurrent 0.509225 0.228524 2.228 0.02586 * smokeFormer 0.408432 0.195298 2.091 0.03650 * waist 0.003483 0.007009 0.497 0.61920 diabYes 0.483286 0.148545 3.253 0.00114 ** hypercholYes -0.323312 0.146776 -2.203 0.02761 * p14 -0.110255 0.035236 -3.129 0.00175 ** --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1</p> <p>(Dispersion parameter for binomial family taken to be 1)</p> <p>Null deviance: 1885.9 on 5660 degrees of freedom Residual deviance: 1764.4 on 5650 degrees of freedom AIC: 1786.4</p> <p>Number of Fisher Scoring iterations: 6</p> <p><i>Il·lustració 5: Resultat regressió logística base d</i></p>

L'il·lustració 4 i 5 mostra la diferència entre dues bases de dades. Es pot observar doncs que al aplicar tota la base de dades (d), que inclou els individus que no contenen

resultats de totes elles, els resultats són quasi idèntics. Per tant, el model és un bon predictor també per la base de dades predimed (de predimed s'extreu la base de dades d amb l'edat desglosada) i no només per la dd. La base dd, no te en compte les dades perdudes ja que s'ha aplicat el filtre na.omit, en canvi en la base d aquesta instrucció no s'ha aplicat. D'aquesta manera podem corroborar que els dos models son correctes i que donen dades similars.



Il·lustració 6: Representació dels models

La il·lustració 6 descriu que el millor model de selecció de les variables és el model 2 que conté les variables del grup de la dieta mediterrània suplementada amb nous o AOVE, sexe, edat, factor de fumar, la circumferència de la cintura (waist), patir diabetis (diabYes), patir hipercolesterolèmia (hypercholYes) i l'adherència a la dieta mediterrània (p14); ja que el número BIC és el més baix respecte els altres models descrits, i indica que la dispersió de les dades estan més ajustades.

Altrament, la il·lustració 6 suporta les conclusions extretes de la il·lustració 1, on s'indicava que la dieta mediterrània amb suplementació de nous és un factor significatiu, en canvi, la dieta mediterrània amb suplementació d'AOVE, no indica tenir cap relació significativa.

		2.5 %	97.5 %	
(Intercept)	0.00	0.00	0.01	Finalment per acabar d'analitzar
sexFemale	0.59	0.41	0.85	el Model, s'ha buscat l'odds i
groupMedDiet + Nuts	0.66	0.46	0.92	l'interval de confiança. A la
groupMedDiet + VOO	0.78	0.57	1.07	il·lustració 7 es pot observar que
age	1.08	1.05	1.10	seguir una dieta suplementada
smokeCurrent	1.66	1.06	2.59	amb nous, sexFemale,
smokeFormer	1.50	1.03	2.21	hyperchol i p14 disminuirien la
waist	1.00	0.99	1.02	probabilitat de què succeeixi
diabYes	1.62	1.21	2.18	una complicació cardiovascular,
hypercholYes	0.72	0.54	0.97	tot i que tant en hypercholYes i
p14	0.90	0.84	0.96	

Il·lustració 7: Odds ratio i IC model 2

p14 donen valors de IC molt proper a 1. A més, l'age, smokeCurrent, smokeFormer i diabYes són variables a tenir en compte com a factors de risc, ja que la seva odds és més alta de 1. Altrament, la dieta suplementada amb AOVE i el waist contenen 1 a l'interval de confiança, indicant que no es pot establir de forma significativa que tinguin algun efectes sobre el fet de patir complicacions cardiovasculars.

Al realitzar el càlcul d'odds ratio i interval de confiança s'ha agafat el model 2 com a millor indicador. Ja que aquest no conte el wth, aquesta variable al fer el càlcul de la odds ens donava uns valors molt més elevats d'1, de manera que em considerat que en relació amb la resta de variables no te sentit un valor tant elevat l'hem de descartar. A més wth es la relació entre cintura i altura, el que la fa una mesura no primària sinó derivada d'ambdues. Es pera quest fet que em fet el càlcul del model tot i que el model que te més significació es el model 3.

Per calcular la probabilitat en el model s'ha aplicat type='response', d'aquesta manera s'ha observat la probabilitat en base al model2 de partir una complicació cardiovascular, amb només la diferència del sexe entre els dos individus, es pot observar doncs que la difèrenia entre homes i dones es de 0,02 de patir al malaltia amb la resta de variables iguals(il·lustració 8).

	sex	smoke	group	hyperchol	age	waist	diab	p14
1	Female	Current	MedDiet + Nuts	Yes	60	120	Yes	10
2	Male	Current	MedDiet + Nuts	Yes	60	120	Yes	10
					1	2		
					0.01927611	0.03233995		

Il·lustració 8: Diferència entre dones i homes de patir l'event amb les mateixes variables.

Anàlisi de supervivència:

```
Call:
coxph(formula = predimed.surv ~ group, data = df)

n= 6324, number of events= 252

              coef exp(coef) se(coef)      z Pr(>|z|)
groupMedDiet + Nuts -0.4190   0.6577  0.1569 -2.671 0.0081
groupMedDiet + VOO  -0.3526   0.7028  0.1487 -2.371 0.0181
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

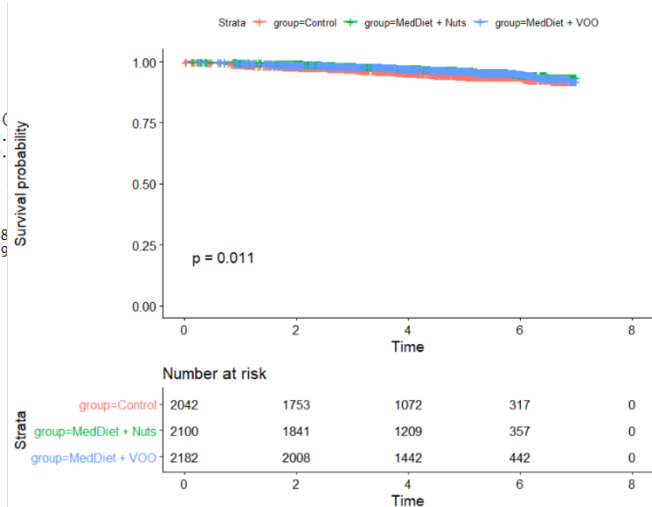
              exp(coef) exp(-coef) lower .95 upper .95
groupMedDiet + Nuts   0.6577     1.520   0.4836   0.898
groupMedDiet + VOO   0.7028     1.423   0.5251   0.942

Concordance= 0.559 (se = 0.018 )
Likelihood ratio test= 8.62 on 2 df,  p=0.01
Wald test               = 8.89 on 2 df,  p=0.01
Score (logrank) test = 9 on 2 df,  p=0.01
```

Il·lustració 10: Significació

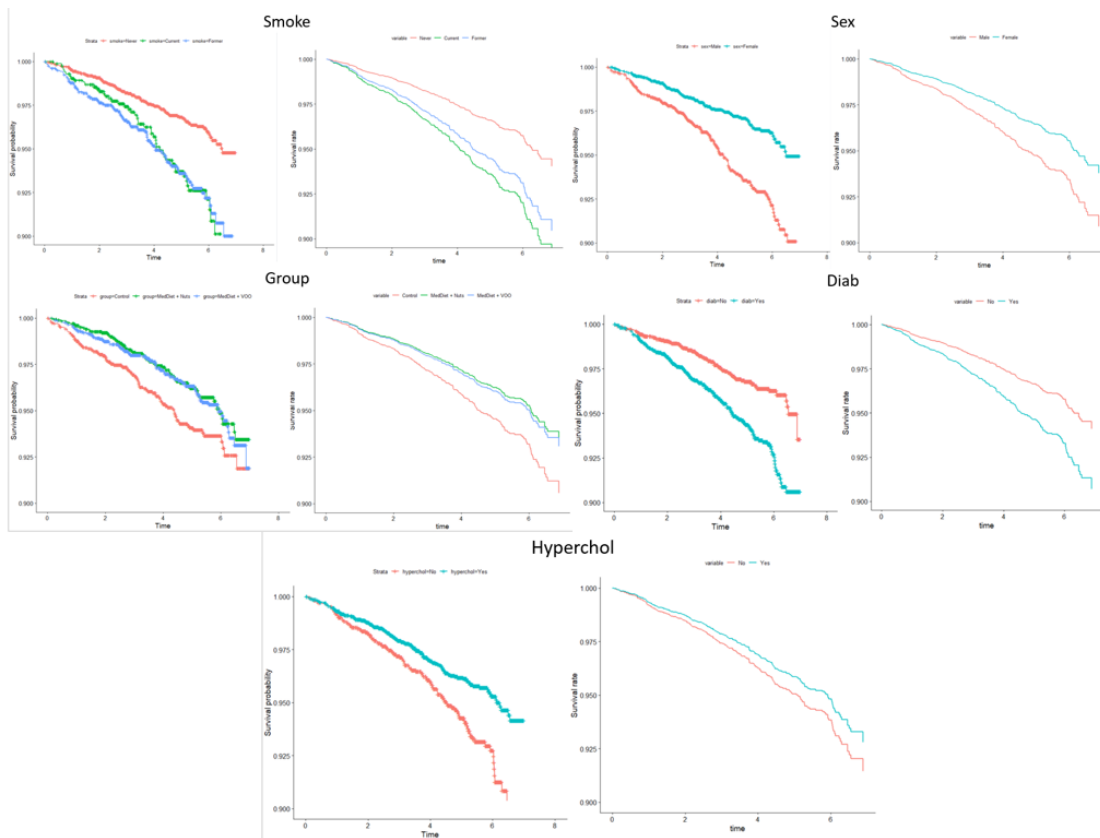
La il·lustració 9 compara les corbes de supervivència en funció de la dieta mediterrània. El valor

$p=0.011$ és significatiu i demostra que la supervivència s'allunya del concepte que no hi ha diferència entre les dietes. Altrament la corba mostra com els valors són molts propers a 1, i en $t=0$ la supervivència és absoluta.



Il·lustració 9: Anàlisi de cox en funció de l'aparició de l'event i el grup

Supportant la il·lustració 9, la il·lustració 10 mostra com tant la dieta suplementada amb AOVE com la dieta suplementada amb nous són significatius, i a més, els IC són menors de 1, per tant, el tractament amb aquestes dietes redueix el risc de tenir complicacions cardiovasculars. Tot i que, el tractament amb la dieta suplementada amb nous és més eficaç en la reducció del risc de patir complicacions cardiovasculars.

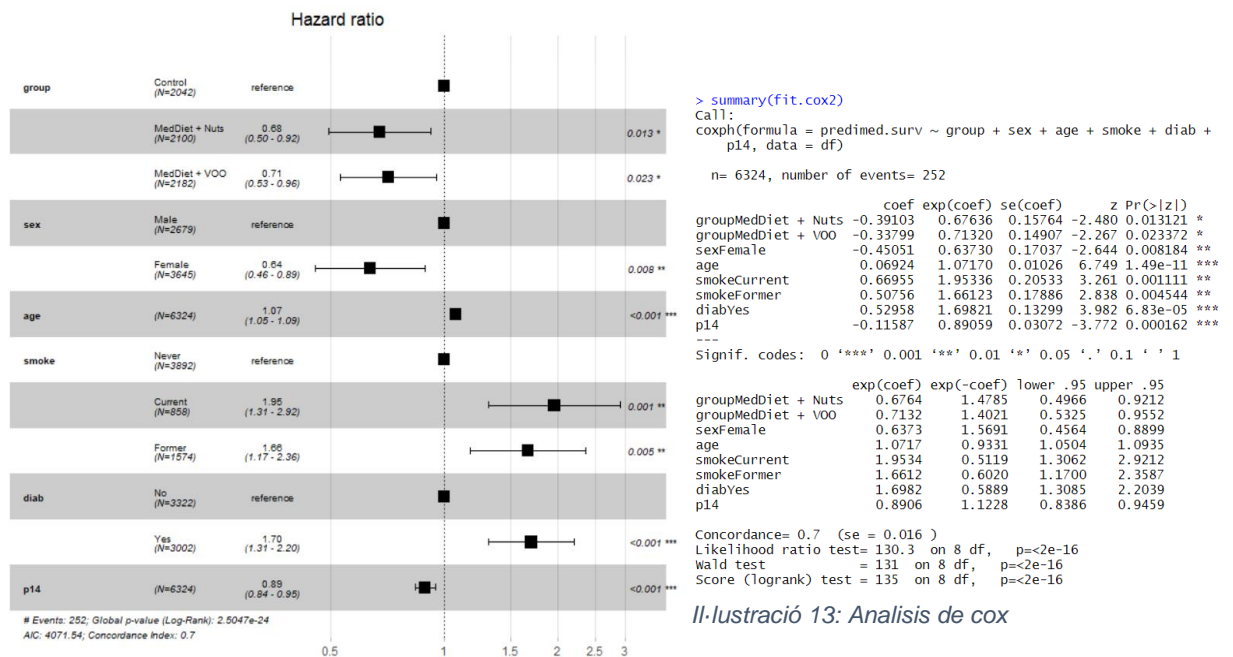


Il·lustració 11: Representació gràfica de les diferents variables nominals

Les variables descrites a la il·lustració 11 que tenen més significació són el factor de fumar, la dieta mediterrània, el sexe, la diabetis, la circumferència de la cintura i l'adherència a la dieta mediterrània. Les corbes de supervivència mostren que no fumar, el sexe femení, la dieta mediterrània suplementada amb nous i AOVE, no patir diabetis i tenir hipercolesterolèmia són factors que augmenten la supervivència, i per tant, poden retardar l'aparició de patir complicacions cardiovasculars.

Les variables edat, el factor de fumar i la diabetis són tenen un IC major d'1, per tant, es poden considerar factors de risc davant les complicacions cardiovasculars; en canvi, la dieta mediterrània, l'adherència a la dieta mediterrània i el sexe femení tenen un IC menor d'1, doncs poden ser considerades factors protectors davant el fet de patir complicacions cardiovasculars.

Conclusions:



Il·lustració 13: Anàlisi de cox

Il·lustració 12: Hazard ratio de les variables significatives

PREDIMED és un estudi que avalua la influència de la dieta mediterrània, suplementada amb nous o AOVE, però no són les úniques variables que tenen relació en la reducció de patir complicacions cardiovasculars.

Els resultats mostren que les dieta mediterrània suplementada amb nous i AOVE influeixen de forma significativa en la reducció de patir complicacions cardiovasculars, essent la dieta mediterrània suplementada amb nous la que mostra una major significació en la reducció de patir complicacions cardiovasculars. Podem concloure doncs que el nostre anàlisi concorda amb les conclusions extretes pels experts en l'estudi, on aquest es destacava que tant la dieta mediterrània suplementada amb AOVE o nous tenen una menor taxa d'esdeveniments cardiovasculars. Corroborant aquesta hipòtesis també està representat en la il·lustració 11 on s'observa que entre els grups de dieta mediterrània amb suplementos de nous i AOVE la diferència entre la probabilitat de supervivència es quasi nul·la, en canvi la diferència entre aquestes dos dietes i la control es d'un 0'020 aproximadament.

A més a més, les variables descrites a l'estudi que també són significatives són el sexe, edat, el factor de fumar, la circumferència de la cintura, el ratio cintura-alçada, patir diabetis i hipercolesterolèmia i l'adherència a la dieta mediterrània, essent el sexe femení, la hipercolesterolèmia i l'adherència a la dieta mediterrània variables

significatives que afavoreixen la reducció de que succeeixi o que apareixi una complicació cardiovascular. De manera que podem concloure dient que fumar, l'edat, el sexe masculí, una major circumferència de cintura, com més gran es a ratio cintura-alçada, tenir diabetis, hipercolesterolèmia son factors de risc per acabar desenvolupant una complicació cardiovascular al llarg de la vida.

Referenciant la il·lustració 7, aquesta ens corrobora aquestes conclusions ja que podem observar com, l'edat, ser fumador actualment o en el passat i tenir diabetis presenten una odds ratio major a 1, el que significa que són factors de risc que podrien derivar en una complicació cardiovascular.

Cal destacar també la diferència que hi ha entre el sexe masculí i femení en presentar una complicació cardiovascular al llarg de la seva vida. Com es pot observar en la il·lustració 11, entre temps 6 i 8 hi ha una diferència de 0.025 aproximadament entre la probabilitat de supervivència entre aquests dos sexes.

Finalment ens va cridar molt l'atenció que les variables bmi i la història familiar no fossin significatives i destacar també el resultat extremadament elevat de wth que ens va fer descartar-lo del nostre anàlisis per les incongruències que