



UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

Trabajo Final

“Calidad de vida de las personas”

PROFESOR: RODRIGO ORTEGA

AYUDANTE: DIEGO BASCUÑAN

INTEGRANTE: BARBARA LIZAMA



Econometría

Contenido

1. Introducción	3
2. Hipótesis del trabajo	4
3. Objetivos.....	4
3.1 Objetivo general.....	4
3.2 Objetivo específico	4
4. Materiales y métodos.....	4
5. Estadística descriptiva	4
6. Planteamiento del Modelo.....	9
6.1 Significancia Global del Modelo original.....	9
6.2 Significancia Individual de Parámetros	10
6.3 Bondad de Ajuste del Modelo	11
7. Modelo 2.....	11
8. VIF (Multicolinealidad)	11
9. Gráficos de Residuales.....	12
10. Análisis Heterocedasticidad	12
11. Corrección Heterocedasticidad.....	13
12. Modelo propuesto	13
13. Bibliografía	14

1. Introducción.

Desde hace varios años todas las personas se encuentran inmersas en un mundo más globalizado, donde todos los países buscan poder optar a ser mejor por vías al desarrollo, concentrando todo el interés en poder crecer más económicamente y poder ser estables dentro de todos los cambios que han ocurrido, ocurren y ocurrirán.

Hoy en día el apostar a estar dentro de este círculo y cumplir las expectativas interpuestas en el mercado, ha comenzado a repercutir en las personas más de lo esperado. El bienestar y la calidad de vida ha sido un factor determinante en este proceso de cambios, considerando que la calidad de vida es un concepto que hace alusión a varios niveles de generalización pasando por sociedad, comunidad, hasta el aspecto físico y mental.

La calidad de vida se puede ver afectadas por muchos factores como, por ejemplo:

1. Físico: donde puede tener aspectos de salud y seguridad física.
2. Material: contar con una vivienda digna y propia, con acceso a todos los servicios básicos, alimentos al alcance, medios de transporte.
3. Educativos: existencia y cercanía a una educación donde permita ampliar los conocimientos de forma individual y colectiva.
4. Emocional: campo muy importante para cualquier ser humano, mediante el cual permite que su estado emocional vaya creciendo e interrelacionarse eficazmente.
5. Social: donde al ser humano tiene opciones interpersonales con núcleos básicos como la familia y círculos de amistades que permitan desarrollo del mismo.

En el mundo estos aspectos han tomado gran relevancia y para que cada país pueda ser mejor debe preocuparse más del bienestar y calidad de vida de sus integrantes, específicamente porque no se puede avanzar a ser más si las personas que componen una nación no tienen una buena condición de salud, física y emocional.



Ilustración 1 respiralibre.cl



Econometría

2. Hipótesis del trabajo

El IMC se ve afectado de manera directa por el estilo de vida considerando para tal fin, la alimentación y niveles de estrés, como también la edad de los individuos considerados para este estudio.

3. Objetivos

3.1 Objetivo general

El objetivo general es determinar que existe un cambio en el IMC y la calidad de vida de las personas a través del bienestar, considerando que las personas se han visto afectadas por los cambios del entorno.

3.2 Objetivo específico

- Encontrar que elementos afectan el bienestar y la calidad de vida de las personas.
- Definir cuál o cuáles variables son relevantes en las personas.

4. Materiales y métodos

El análisis del bienestar y calidad de vida se realizó con la base de datos obtenida de www.kaggle.com, sitio en el cual se puede obtener el resultado de encuestas realizadas en diferentes ciudades de EE. UU, lugares y personas en un periodo determinado.

El periodo por analizar abarca datos desde el año 2015 al año 2020, la encuesta se realizó a 5.877 individuos.

Para realizar el trabajo se seleccionaron 5 variables.

Definición de variables:

- IMC
- Edad
- Género
- Estrés diario
- Fruta y verdura

En cuanto al software utilizado, estos corresponden a Excel con su respectivo Complemento de Análisis de Datos y Stata 16.

5. Estadística descriptiva.

Se presenta un análisis descriptivo de la encuesta titulada “Bienestar y estilo de vida” que aplica las siguientes variables:

- IMC: índice de masa corporal
- Edad: rangos de edad de los encuestados
- Género: sexo del encuestado



Econometría

- Estrés diario: nivel de estrés del encuestado, evaluado de 0 a 5.

Nivel	
0	Sin Estrés
1	Estrés Leve
2	Estrés Moderado
3	Estrés Severo
4	Estresado
5	Muy Estresado

Ilustración 2 Descripción Niveles de Estrés

- Fruta y verdura: consumo de frutas o verduras del encuestado, de 0 a 5.

Frutas o Verduras	
0	Sin consumo
1	Consume 1 fruta o verdura diaria
2	Consume 2 frutas o verduras diarias
3	Consume 3 frutas o verduras diarias
4	Consume 4 frutas o verduras diarias
5	Consume 5 frutas o verduras diarias

Ilustración 3 Descripción Niveles variable frutas o verduras

La encuesta fue aplicada durante el periodo de 2015 a 2020 a 5.877 individuos, donde el mayor número de encuestados fue del género femenino y en un rango etario de 21 a 35 años de edad.

En promedio el consumo de frutas y/o verduras diarias de un individuo es de 2,9 unidades, el promedio de índice de masa corporal entre los individuos es de 1,4 aprox lo cual no es tan malo considerando que 1 es buen IMC Y 2 mal IMC.

	Frutas Verduras	Estrés Diario	Rango IMC	Edad	Genero
Media	2,9275	2,7791	1,4048	1,6052	0,6141
Error típico	0,0189	0,0180	0,0064	0,0124	0,0064
Mediana	3	3	1	2	1
Moda	3	3	1	1	1
Desviación estándar	1,4457	1,3786	0,4909	0,9505	0,4869
Varianza de la muestra	2,0900	1,9005	0,2410	0,9034	0,2370
Curtosis	-1,0150	-0,8154	-1,8501	-0,9761	-1,7808
Coefficiente de asimetría	-0,0674	-0,0587	0,3880	0,0337	-0,4688
Rango	5	5	1	3	1
Mínimo	0	0	1	0	0
Máximo	5	5	2	3	1
Suma	17.205	16.333	8.256	9.434	3.609
Cuenta	5.877	5.877	5.877	5.877	5.877
Nivel de confianza(95,0%)	0,0370	0,0353	0,0126	0,0243	0,0124

Ilustración 4 Estadística Descriptiva

Para un mejor entendimiento de la tabla anterior, es necesario especificar que las variables Edad y Género fueron categorizadas para facilitar su análisis.



Econometría

Para la variable Genero: Mujer=1; Hombre=0 y para la variable Edad: 1- (21 to 35), 2-(36 to 50), 3- (51 or more), 0 – (less than 20).

De acuerdo a la especificación anterior, la mayor parte de los encuestados tienen entre 21 y 35 años.

Test de Shapiro Wilk

Respecto a las variables, se realizó un test de Shapiro Wilk, que contrasta las siguientes hipótesis:

H_0 : las variables provienen de una distribución normal VS H_a : No provienen de una distribución normal

Se han obtenidos los siguientes datos:

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
edad_cat	5,877	0.99535	14.551	7.061	0.00000
gen_bin	5,877	0.99984	0.516	-1.746	0.95958
RangoIMC	5,877	0.99987	0.392	-2.469	0.99322
EstrésDiario	5,877	0.99785	6.738	5.031	0.00000
FrutasVerd~s	5,877	0.99468	16.649	7.416	0.00000
Año	5,877	0.98462	48.102	10.213	0.00000

Note: The normal approximation to the sampling distribution of W' is valid for $4 \leq n \leq 2000$.

Ilustración 5 Test Shapiro Wilk

Según los resultados anteriores no todas las variables tienen una distribución normal, sin embargo, en casi todos los casos no se rechaza H_0 , pero se debe tener en cuenta que esta prueba es estricta en el número de observaciones y en este caso, la encuesta analizada supera las 2.000 observaciones.

Análisis grafico

Mediante las gráficas siguientes, se puede plantear que, mediante el aumento del número de observaciones, la tendencia de estas seguirá una distribución aproximadamente normal.

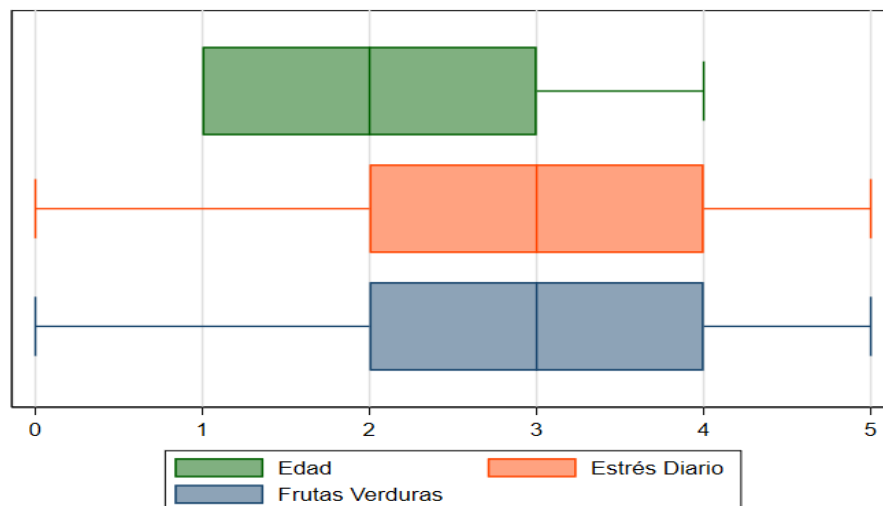


Ilustración 6 Grafico variables



Econometría

Además, se presentan las siguientes graficas de barra para mejorar la comprensión de este análisis, para ello se presenta la frecuencia del variable sexo, y la frecuencia de estrés diario segmentada por rango etario.

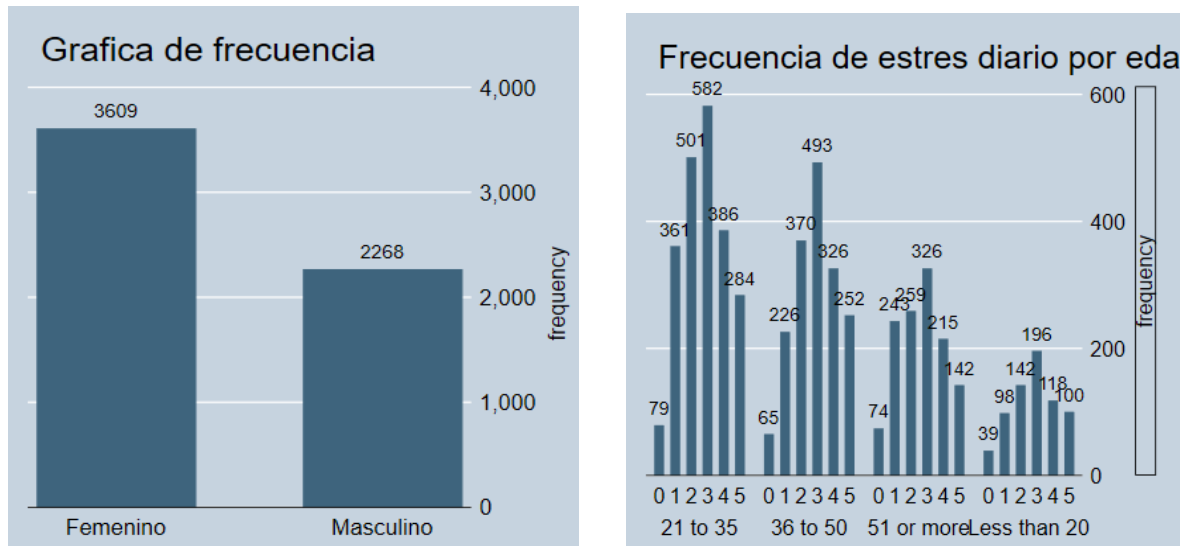


Ilustración 7 Graficas de barra

Como fue mencionado anteriormente, la mayor cantidad de encuestado fue mujeres. De acuerdo al grafico de frecuencia de estrés diario por edad, es notable mencionar que el rango etario con mayores niveles de estrés es el de 21 a 35 años de edad, donde la mayoría de los individuos con un número igual a 582 respondió tener un nivel de estrés igual a 3 que representa un estrés severo. Cabe destacar que solo 39 individuos menores de 20 años consideran no tener o sentir estrés.

Mediante este análisis surge la pregunta, ¿el género de un individuo tiene relación con el nivel de estrés que este siente? Para responder a esta pregunta se procede a generar una tabla cruzada y hacer un test Chi- cuadrado.

		Nivel de estrés						Total
		0	1	2	3	4	5	
Genero	0	140	443	527	547	374	237	2.268
	1	117	485	745	1050	671	541	3.609
	Total	257	928	1.272	1.597	1.045	778	5.877
		4%	16%	22%	27%	18%	13%	

		Nivel de estrés						Total
		0	1	2	3	4	5	
Genero	0	99,2	358,1	490,9	616,3	403,3	300,2	2.268
	1	157,8	569,9	781,1	980,7	641,7	477,8	3.609
	Total	257	928	1272	1597	1045	778	5.877

Distancia de Chi- cuadrado						
16,8013	20,1149	2,6579	7,7925	2,1255	13,3199	
10,5584	12,6408	1,6703	4,8970	1,3357	8,3706	

chi-cal	102,2850
chi-tabla	9,4877

Ilustración 8 Estadístico Chi-cuadrado



Econometría

Como el Chi-calculado es mayor al Chi de tabla podemos decir que existe una relación estadísticamente significativa entre las variables género y nivel de estrés de los individuos. Que en este caso implica que el hecho de ser mujer aumenta las probabilidades de sentir estrés.

6. Planteamiento del Modelo

Modelo Original

IMC= B0 + B1*Frutas Verduras + B2*Estrés Diario + B3*Gen +B4*Edad.

Regression Analysis

OVERALL FIT

Multiple R	0,2555	AIC	-8.750,9170
R Square	0,0653	AICc	-8.750,9027
Adjusted R Square	0,0646	SBC	-8.717,5230
Standard Error	0,4748		
Observations	5877		

ANOVA				Alpha	0,05	
	df	SS	MS	F	p-value	sig
Regression	4	92,4129	23,1032	102,4970	1,68824E-84	yes
Residual	5872	1.323,5718	0,2254			
Total	5876	1.415,9847				

	coeff	std err	t stat	p-value	lower	upper	vif
Intercept	1,2464	0,0218	57,1644	-	1,20364	1,28913	
Frutas y Verduras	-0,0416	0,0044	-9,4684	4,0180,E-21	-0,05016	-0,03295	1,04969
Estrés Diario	0,0374	0,0046	8,2054	2,8002,E-16	0,02844	0,04630	1,02771
Genero	-0,0065	0,0129	-0,5060	0,6129	-0,03191	0,01882	1,03412
Edad	0,1123	0,0066	16,9411	6,8447,E-63	0,09929	0,12527	1,03446

Ilustración 9 Modelo Regresión Múltiple Real Statistics

6.1 Significancia Global del Modelo original

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$

$$H_1: \text{Algun } \beta_i \neq 0$$

El valor del estadístico F es:

$$F(4, 5872) = 102,4970$$

Cuyo p- value es:

$$P\text{-value} = 1,6882E-84$$

Se puede concluir que se rechaza la hipótesis nula debido a que el p-value es menor al alpha, por lo cual el modelo es significativo globalmente.



Econometría

6.2 Significancia Individual de Parámetros

Para el coeficiente B1: (Frutas y verduras)

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

P value = 4,0108E-21

Dado que p value < alpha, se rechaza la hipótesis nula, por lo cual la variable es significativa localmente

Para el coeficiente B2: (Estrés Diario)

$$H_0: \beta_2 = 0$$

$$H_1: \beta_2 \neq 0$$

P value = 2,8002E-16

Dado que p value < alpha, se rechaza la hipótesis nula, por lo cual la variable es significativa localmente.

Para el coeficiente B3: (Género)

$$H_0: \beta_3 = 0$$

$$H_1: \beta_3 \neq 0$$

P value = 0,6129

Dado que p value > alpha, no se rechaza la hipótesis nula, por lo cual la variable no es significativa localmente.

Para el coeficiente B4: (Edad)

$$H_0: \beta_4 = 0$$

$$H_1: \beta_4 \neq 0$$

P value = 6,8447E-63

Dado que p valor < alpha, se rechaza la hipótesis nula, por lo cual la variable es significativa localmente.

Dados los análisis de significancia local individual se realizará eliminación del variable género por no resultar significativa localmente en el modelo.



Econometría

6.3 Bondad de Ajuste del Modelo

R-squared = 0,0653

El valor de R^2 indica que el 6,52% de las variaciones en el índice de masa corporal se explican por el modelo presentado.

7. Modelo 2

Se realiza eliminando la variable Género

Regression Analysis

OVERALL FIT

Multiple R	0,2554	AIC	-8752,660831
R Square	0,0652	AICc	-8752,650611
Adjusted R Square	0,0647	SBC	-8725,945624
Standard Error	0,4747		
Observations	5.877		

ANOVA

	df	SS	MS	F	p-value	sig
Regression	3	92,3552	30,7851	136,5946	1,5079E-85	yes
Residual	5.873	1323,6295	0,2254			
Total	5.876	1415,9847				

	coeff	std err	t stat	p-value	lower	upper	vif
Intercept	1,2443	0,0214	58,1478	0	1,2023	1,2862	
Frutas y Verduras	-0,0418	0,0044	-9,5743	1,47238E-21	-0,0503	-0,0332	1,0382
Estrés Diario	0,0371	0,0045	8,2126	2,63711E-16	0,0282	0,0459	1,0090
Edad	0,1121	0,0066	16,9467	6,24879E-63	0,0991	0,1250	1,0297

Ilustración 10 Modelo 2 Regresión Múltiple

En este caso, se obtiene un modelo significativo en forma global, con todos sus parámetros significativos en forma individual.

8. VIF (Multicolinealidad)

	vif
Frutas y Verduras	1,04969
Estrés Diario	1,02771
Genero	1,03412
Edad	1,03446

Ilustración 11 VIF Modelo Original

Dado que los VIF son todos menores a 10, entonces no existe problema de multicolinealidad.



Econometría

9. Gráficos de Residuales

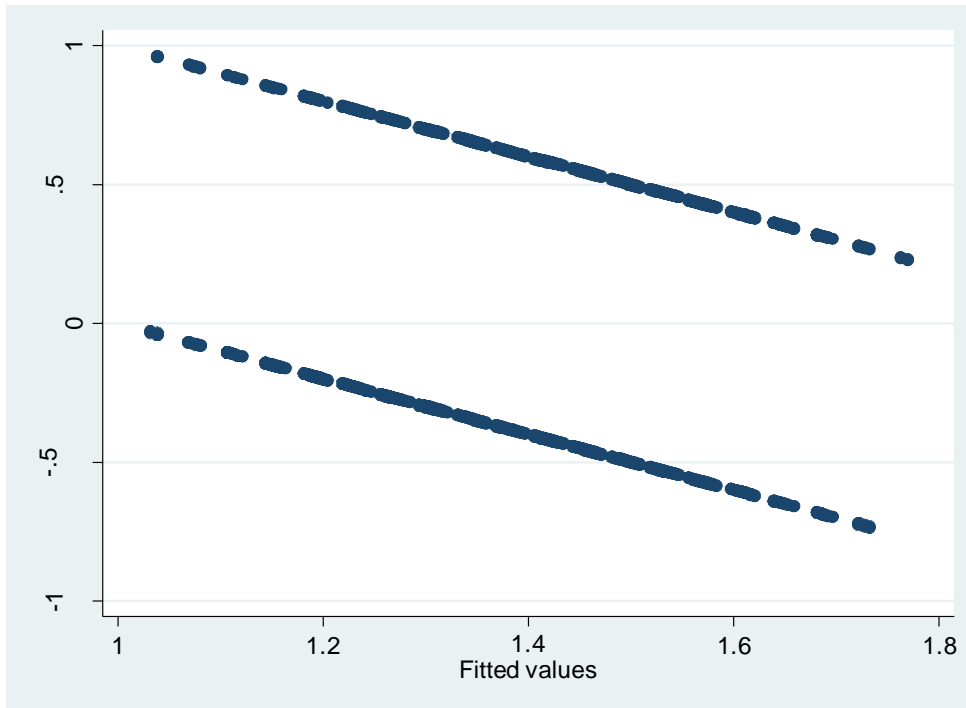


Ilustración 12 Grafico Residuales vs Valores ajustados

10. Análisis Heterocedasticidad

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
Ho: Constant variance
Variables: fitted values of bmi_range

chi2(1)      =    26.54
Prob > chi2  =    0.0000
```

Ilustración 13 Prueba de Breusch-Pagan

Dado que p valor es menor a 5%, entonces se rechaza la hipótesis nula de varianza constante, es decir hay problema de heterocedasticidad.

Observaciones influyentes (residuales estudentizados, leverage values, DFFITS, DFBETAS, COOK's distance, Covratio)

```
. predict cook, cooks
```

```
. count if cook>1 & cook<.0
```

De acuerdo con la distancia de Cook, no habría observaciones influyentes.



Econometría

11. Corrección Heterocedasticidad

Se realiza con MCG, mínimos cuadrados Generalizados

El modelo finalmente obtenido es significativo en forma global, con significancia individual de parámetros, y con heterocedasticidad corregida.

```
. glm imc frutasyverduras EstrésDiario edad, family(gaussian)link(identity)

Iteration 0:   log likelihood = -3958.7713

Generalized linear models                               No. of obs   =       5,877
Optimization      : ML                               Residual df   =       5,873
                                                         Scale parameter =    .2253754
Deviance          = 1323.629491                       (1/df) Deviance =    .2253754
Pearson           = 1323.629491                       (1/df) Pearson  =    .2253754

Variance function: V(u) = 1                           [Gaussian]
Link function      : g(u) = u                         [Identity]

Log likelihood    = -3958.771344                       AIC           =    1.348569
                                                         BIC           =   -49646.97
```

imc	OIM		z	P> z	[95% Conf. Interval]	
	Coef.	Std. Err.				
frutasyverduras	-.0417914	.004365	-9.57	0.000	-.0503466	-.0332363
EstrésDiario	.0370606	.0045126	8.21	0.000	.028216	.0459052
edad	.1120551	.0066122	16.95	0.000	.0990954	.1250148
_cons	1.244271	.0213984	58.15	0.000	1.202331	1.286211

12. Modelo propuesto.

$$\text{IMC} = 1,2442 - 0,0418 * \text{frutas y verduras} + 0,0371 * \text{estrés diario} + 0,1121 * \text{edad}$$

El estrés diario contribuye a un aumento de la masa corporal junto con la edad, en tanto que consumir frutas y verduras contribuye a una disminución del IMC.

Se obtiene un modelo significativo con todos sus parámetros calculados significativos, al cual se le realiza corrección por heterocedasticidad mediante mínimos cuadrados generalizados.

La capacidad explicativa del modelo resulta baja, lo cual puede deberse a una mala especificación del modelo por no incluir variables relevantes.



UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

Econometría

13. Bibliografía.

La información para realizar el trabajo propuesto y sus respectivos análisis la he obtenido de las siguientes fuentes:

- Wooldridge, J. (2009). Introducción a la econometría/ Introductory Econometrics: A Modern Approach (4Ta ed.). CENGAGE learning.