

Taller 3 - Modelo Regresión Multinomial

AUTHORS
Jhon Tascon Velasco
Lino Sinisterra
Juan Chacon

PUBLISHED
September 15, 2025

0. Información general

En este trabajo se identificarán qué factores inciden en la preferencia de modo de transporte: avión, tren, bus y carro.

Las variables a analizar son las siguientes:

Descripción de variables del estudio

Variables	Tipo_Variable	Descripción
Tiempo_Espera	Cuantitativa	Tiempo de espera en terminal
Costo_Viaje	Cuantitativa	Costo del viaje
Tiempo_Viaje	Cuantitativa	Tiempo de viaje
Costo_Generalizado	Cuantitativa	Costo generalizado del viaje
Ingresos_Hogar	Cuantitativa	Ingresos del hogar
Tamano_Grupo	Ordinal	Tamaño del grupo que viaja
Modo_Viaje	Nominal/Variable objetivo	Modo de transporte

1. Análisis exploratorio

1.1. Revisión de datos faltantes

Datos faltantes

	x
Modo_Viaje	0
Tiempo_Espera	0
Costo_Viaje	0
Tiempo_Viaje	0
Costo_Generalizado	0
Ingresos_Hogar	0
Tamano_Grupo	0

Con base en la tabla anterior, no hay datos faltantes.

1.2. Resumen de las variables

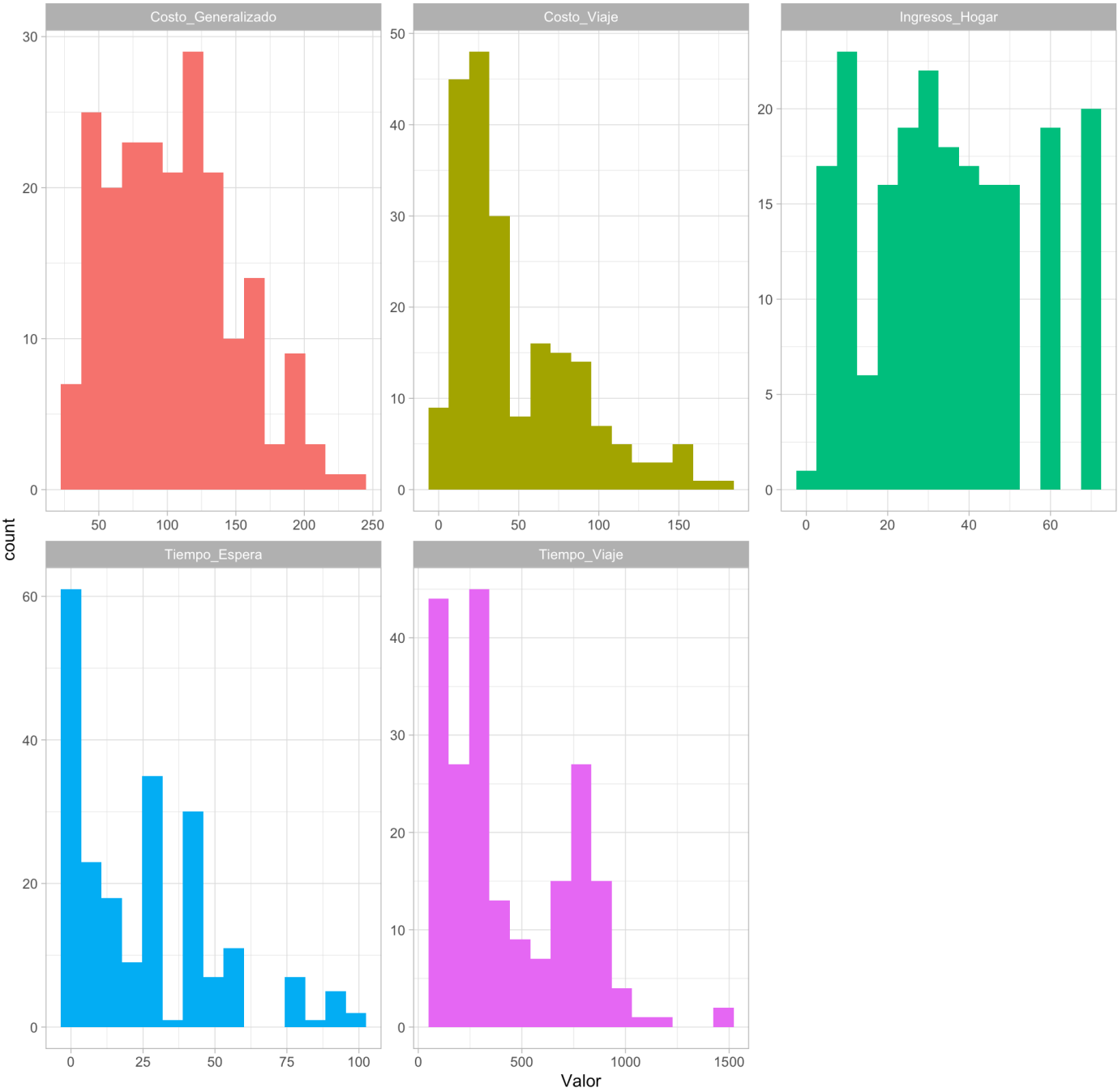
Descripción del conjunto de datos

Modo_Viaje	Tiempo_Espera	Costo_Viaje	Tiempo_Viaje	Costo_Generalizado	Ingresos_Hogar	Tamano_Gru
Avión:58	Min. : 0.00	Min. : 2.0	Min. : 65.0	Min. : 30.0	Min. : 2.00	1:114
Bus :30	1st Qu.: 0.00	1st Qu.: 19.0	1st Qu.: 180.0	1st Qu.: 67.0	1st Qu.:20.00	2: 58
Carro:59	Median :20.00	Median : 33.0	Median : 305.0	Median :102.5	Median :34.50	3: 20
Tren :63	Mean :25.01	Mean : 47.4	Mean : 430.8	Mean :103.8	Mean :34.55	4: 15
NA	3rd Qu.:40.00	3rd Qu.: 70.0	3rd Qu.: 720.0	3rd Qu.:132.8	3rd Qu.:50.00	5: 2
NA	Max. :99.00	Max. :180.0	Max. :1440.0	Max. :238.0	Max. :72.00	6: 1

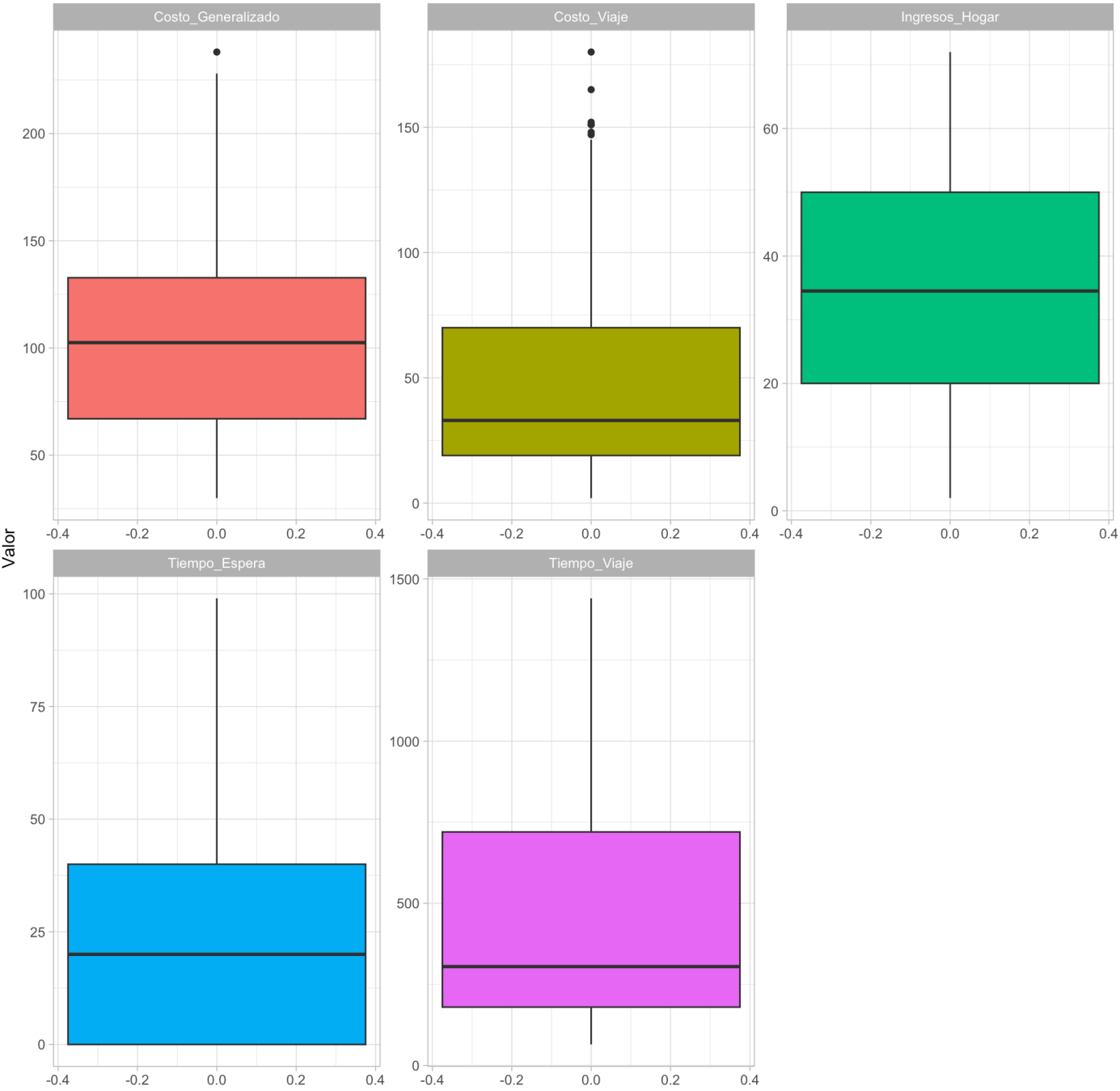


Analizar

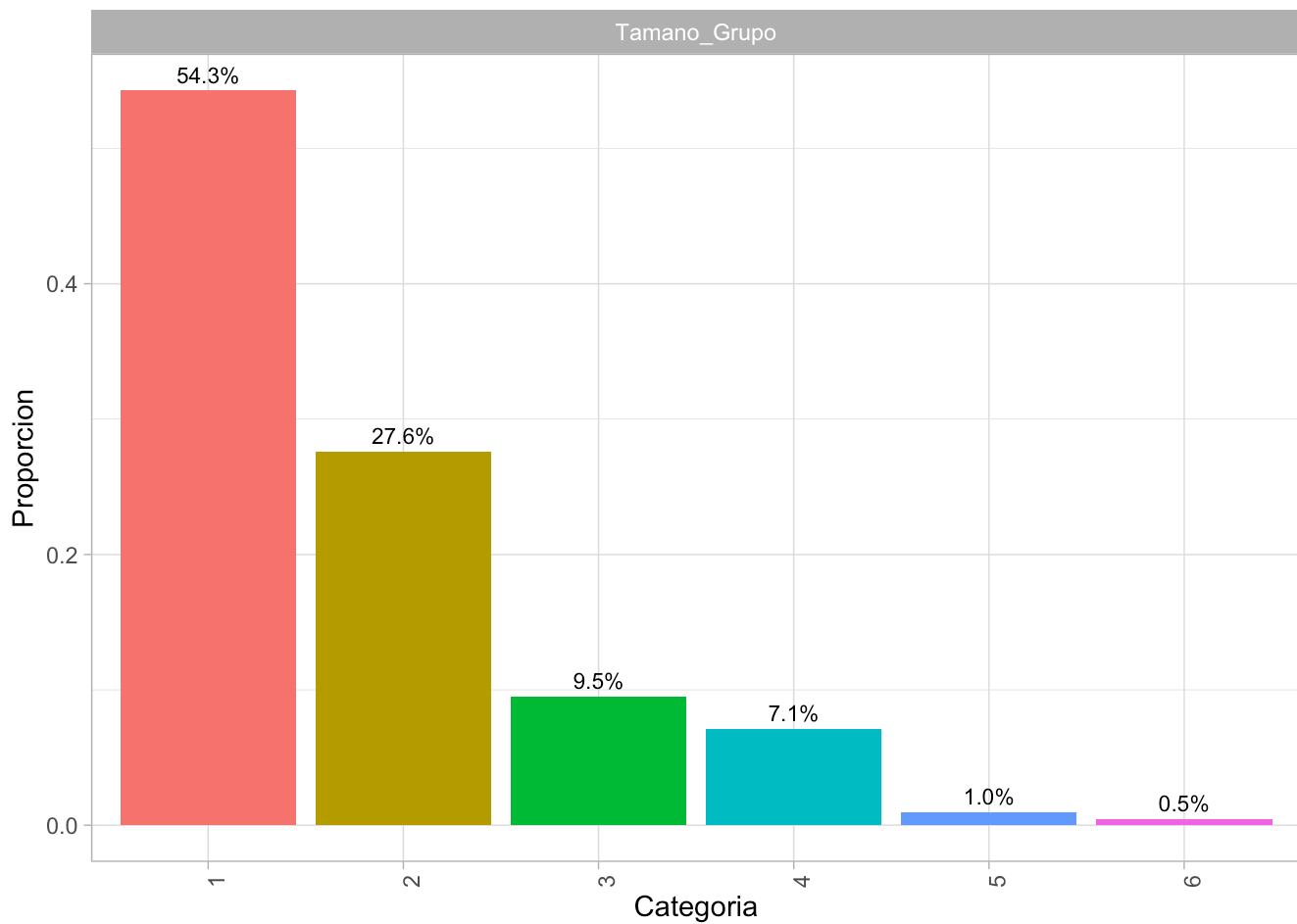
Histogramas de las variables numéricas



Boxplots de las variables numéricas



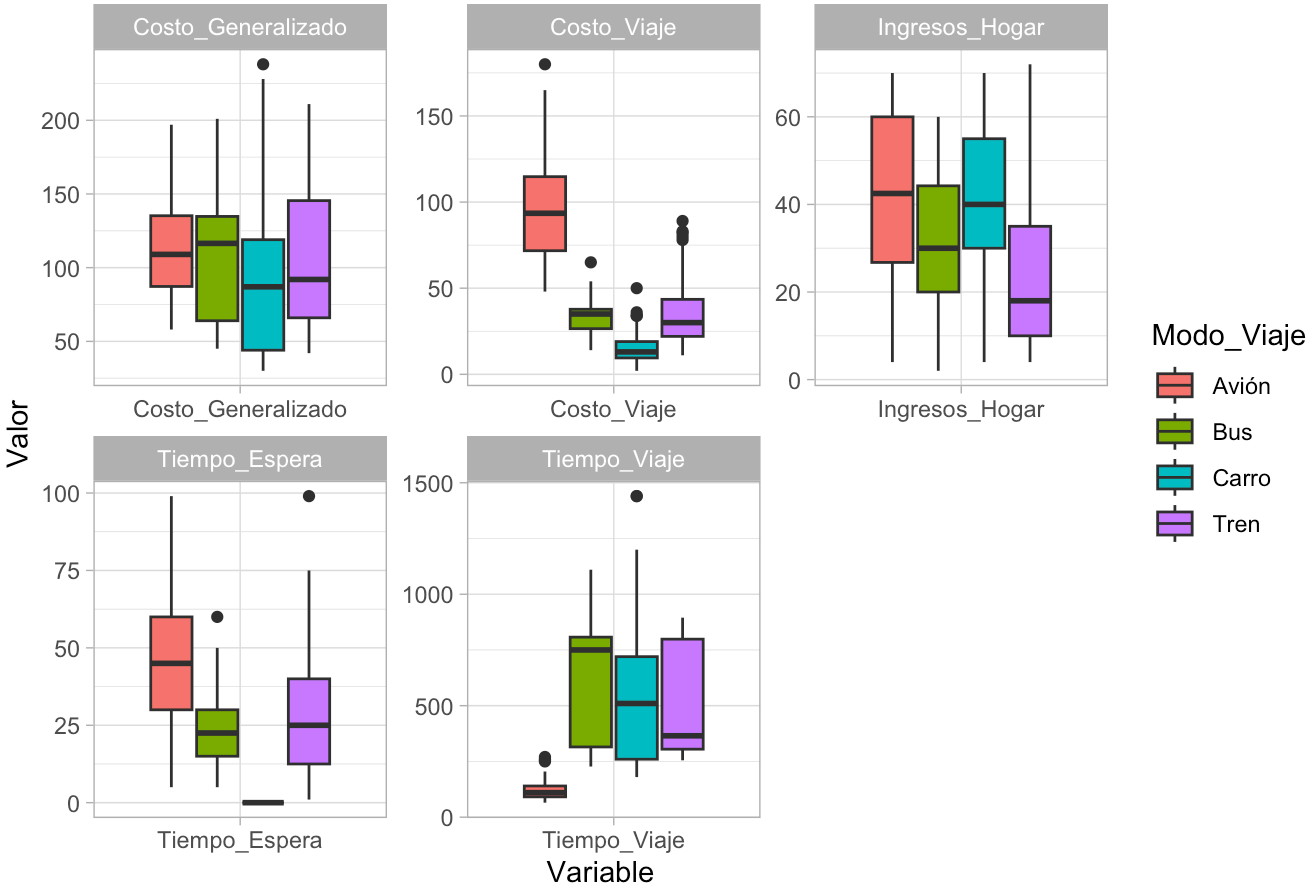
Poner hallazgos



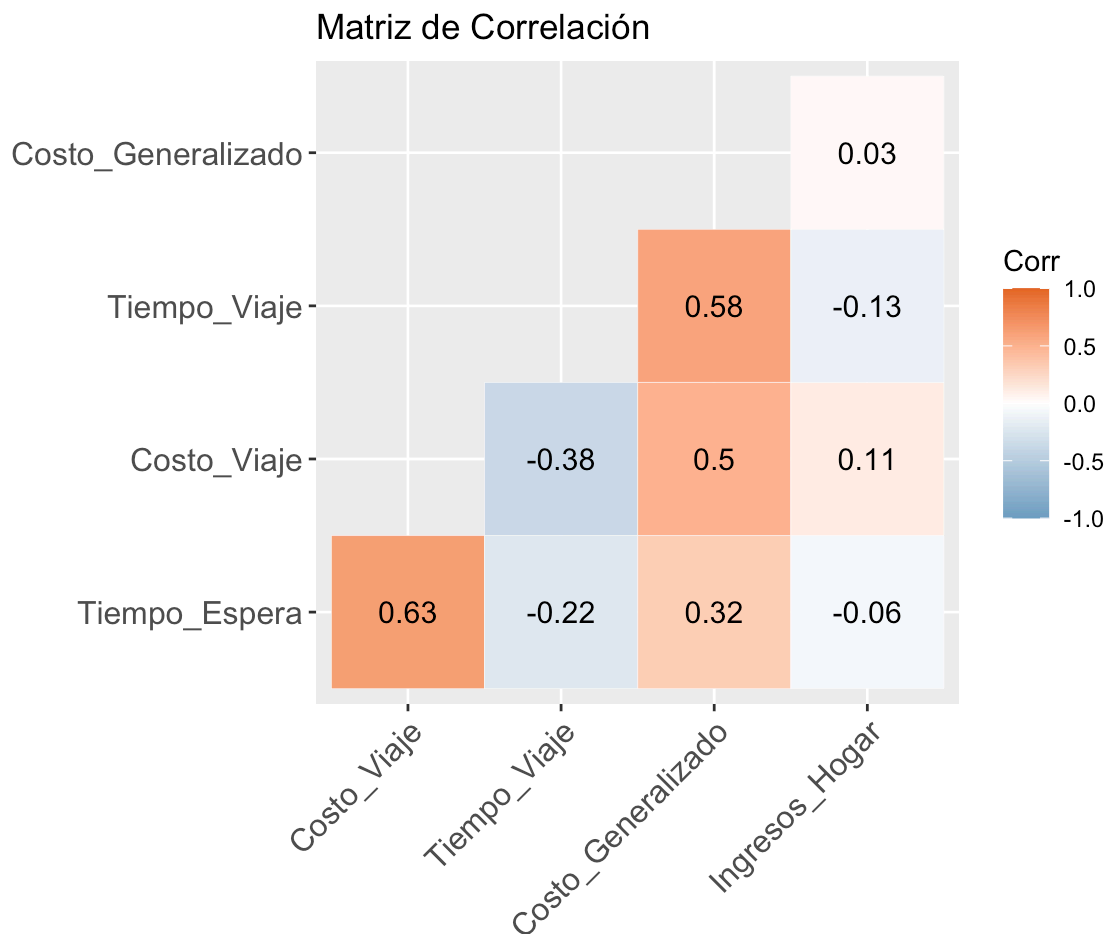
Poner hallazgos

1.3. Análisis bivariado

Distribución de las variables numéricas con el Modo de transporte

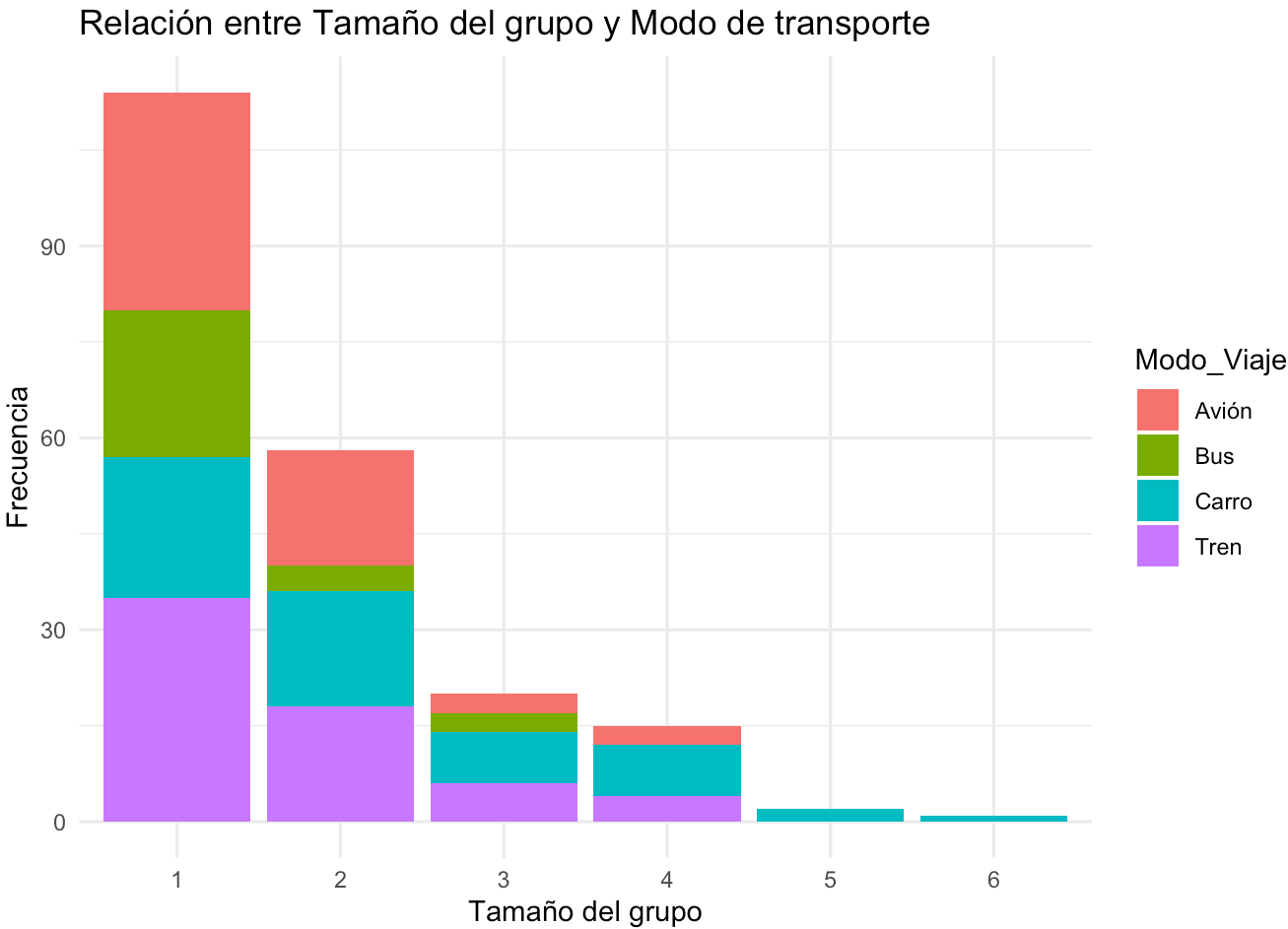


Poner hallazgos



Poner hallazgos

Si bien no se presentan asociaciones muy fuertes entre variables, excepto para tiempo de espera y costo de viaje que tiene una correlación Pearson de 0.63, el Costo generalizado es una combinación lineal del tiempo de transporte y el costo de viaje, por lo tanto, se va a eliminar del análisis para evitar una posible multicolinealidad. Además, revisando las relación entre variables, se evidencia que tiempos de espera en cero corresponden a los carros, por ende, puede sesgar el modelo. Por este motivo, tampoco se tendrá en cuenta dicha variable para el análisis.



Poner hallazgos

2. Modelo multinomial

```
# weights: 40 (27 variable)
initial value 291.121816
iter 10 value 147.782539
iter 20 value 106.601389
iter 30 value 103.150919
iter 40 value 102.805410
iter 50 value 102.803334
final value 102.803331
converged
```

Coefientes

	(Intercept)	Costo_Viaje	Tiempo_Viaje	Ingresos_Hogar	Tamano_Grupo.L
Bus	46.15633	-4.112399	0.5617427	-0.1735036	-88.24744
Carro	78.77012	-4.276600	0.5633502	-0.1358230	-30.83879
Tren	69.46713	-4.081489	0.5594685	-0.2012959	-60.33678
	Tamano_Grupo.Q	Tamano_Grupo.C	Tamano_Grupo^4	Tamano_Grupo^5	
Bus	24.42756	59.00828	-4.453353	-99.21981	

Carro	12.70779	36.21582	21.083529	-48.43903
Tren	-10.30403	25.61379	19.218698	-47.75568

P-values

	(Intercept)	Costo_Viaje	Tiempo_Viaje	Ingresos_Hogar	Tamano_Grupo.L
Bus	0	0	0	0	0
Carro	0	0	0	0	0
Tren	0	0	0	0	0
	Tamano_Grupo.Q	Tamano_Grupo.C	Tamano_Grupo^4	Tamano_Grupo^5	
Bus	0	0	0	0	
Carro	0	0	0	0	
Tren	0	0	0	0	

2.1. Análisis de los coeficientes y p-values

Sale del summary del modelo, tener en cuenta el p-value

2.2. Odds Ratio

	(Intercept)	Costo_Viaje	Tiempo_Viaje	Ingresos_Hogar	Tamano_Grupo.L
Bus	1.110297e+20	0.01636847	1.753726	0.8407142	4.727406e-39
Carro	1.619677e+34	0.01388981	1.756547	0.8729972	4.044674e-14
Tren	1.476355e+30	0.01688231	1.749742	0.8176705	6.252718e-27
	Tamano_Grupo.Q	Tamano_Grupo.C	Tamano_Grupo^4	Tamano_Grupo^5	
Bus	4.062120e+10	4.236145e+25	1.163948e-02	8.116778e-44	
Carro	3.303097e+05	5.349701e+15	1.433706e+09	9.187510e-22	
Tren	3.349776e-05	1.330237e+11	2.221133e+08	1.819588e-21	

Análisis de los odds

2.3. Análisis marginal

Análisis marginal

	term	group	contrast	estimate	std.error	statistic	p.value	s.value	conf.low
	Costo_Viaje	Avión	dY/dX	0.0000000	0.0000000	1.835362e+00	0.0664520	3.9115433	0.0000000
	Costo_Viaje	Bus	dY/dX	0.0034838	0.0013013	2.677195e+00	0.0074242	7.0735579	0.0009330
	Costo_Viaje	Carro	dY/dX	-0.0144705	0.0016538	-8.749793e+00	0.0000000	58.6988235	-0.0177119
	Costo_Viaje	Tren	dY/dX	0.0109867	0.0013141	8.360844e+00	0.0000000	53.8342189	0.0084112
	Ingresos_Hogar	Avión	dY/dX	0.0000000	0.0000000	1.887397e+00	0.0591069	4.0805292	0.0000000
	Ingresos_Hogar	Bus	dY/dX	0.0005479	0.0012306	4.451836e-01	0.6561871	0.6078210	-0.0018640
	Ingresos_Hogar	Carro	dY/dX	0.0042728	0.0010796	3.957651e+00	0.0000757	13.6895314	0.0021568
	Ingresos_Hogar	Tren	dY/dX	-0.0048207	0.0012475	-3.864394e+00	0.0001114	13.1324141	-0.0072650
	Tamano_Grupo	Avión	2 - 1	0.0000000	0.0000000	1.589138e+00	0.1120293	3.1580518	0.0000000
	Tamano_Grupo	Avión	3 - 1	0.0000000	0.0000000	-1.519980e+00	0.1285161	2.9599793	0.0000000

term	group	contrast	estimate	std.error	statistic	p.value	s.value	conf.low
Tamano_grupo	Avión	4 - 1	0.0095238	0.0000000	4.050173e+06	0.0000000	Inf	0.0095238
Tamano_grupo	Avión	5 - 1	0.0095238	0.0000000	2.487501e+07	0.0000000	Inf	0.0095238
Tamano_grupo	Avión	6 - 1	0.0000001	0.0000001	8.948997e-01	0.3708407	1.4311285	-0.0000001
Tamano_grupo	Bus	2 - 1	-0.0947621	0.0480636	-1.971598e+00	0.0486555	4.3612521	-0.188965
Tamano_grupo	Bus	3 - 1	-0.0156164	0.0872240	-1.790381e-01	0.8579078	0.2211055	-0.186572
Tamano_grupo	Bus	4 - 1	-0.1746788	0.0306286	-5.703136e+00	0.0000000	26.3412501	-0.234709
Tamano_grupo	Bus	5 - 1	-0.1746788	0.0306286	-5.703136e+00	0.0000000	26.3412502	-0.234709
Tamano_grupo	Bus	6 - 1	-0.1746788	0.0306286	-5.703136e+00	0.0000000	26.3412501	-0.234709
Tamano_grupo	Carro	2 - 1	0.0174830	0.0499041	3.503322e-01	0.7260894	0.4617810	-0.080327
Tamano_grupo	Carro	3 - 1	-0.0377621	0.0805577	-4.687580e-01	0.6392426	0.6455646	-0.195652
Tamano_grupo	Carro	4 - 1	-0.0022215	0.1142994	-1.943610e-02	0.9844932	0.0225468	-0.226244
Tamano_grupo	Carro	5 - 1	0.4207496	0.0305122	1.378956e+01	0.0000000	141.2844166	0.360946
Tamano_grupo	Carro	6 - 1	0.4466213	0.0255671	1.746863e+01	0.0000000	224.5784558	0.396510
Tamano_grupo	Tren	2 - 1	0.0772791	0.0565179	1.367339e+00	0.1715190	2.5435601	-0.033493
Tamano_grupo	Tren	3 - 1	0.0533785	0.0984097	5.424111e-01	0.5875354	0.7672524	-0.139500
Tamano_grupo	Tren	4 - 1	0.1673765	0.1148847	1.456909e+00	0.1451415	2.7844679	-0.057793
Tamano_grupo	Tren	5 - 1	-0.2555946	0.0324663	-7.872622e+00	0.0000000	48.0328013	-0.319227
Tamano_grupo	Tren	6 - 1	-0.2719425	0.0294215	-9.242992e+00	0.0000000	65.1773414	-0.329607
Tiempo_viaje	Avión	dY/dX	0.0000000	0.0000000	-1.844235e+00	0.0651489	3.9401152	0.000000
Tiempo_viaje	Bus	dY/dX	0.0000942	0.0000926	1.017842e+00	0.3087532	1.6954741	-0.000087
Tiempo_viaje	Carro	dY/dX	0.0002324	0.0000856	2.713883e+00	0.0066500	7.2324364	0.000064
Tiempo_viaje	Tren	dY/dX	-0.0003266	0.0000990	-3.298775e+00	0.0009711	10.0081259	-0.000520

Análisis marginal

2.4. Análisis de las predicciones

Probabilidades medias por cada nivel de tamaño de grupo

```
pp.sim$Tamano_grupo: 1
      Avión      Bus      Carro      Tren
1.231788e-56 1.834502e-01 4.383172e-02 7.727181e-01
-----
pp.sim$Tamano_grupo: 2
      Avión      Bus      Carro      Tren
5.021822e-49 7.040037e-02 4.285723e-02 8.867424e-01
-----
pp.sim$Tamano_grupo: 3
      Avión      Bus      Carro      Tren
```

3.532199e-69 1.401616e-01 3.008804e-02 8.297503e-01

pp.sim\$Tamano_Grupo: 4

Avión	Bus	Carro	Tren
4.451402e-30	3.236775e-41	3.127493e-02	9.687251e-01

pp.sim\$Tamano_Grupo: 5

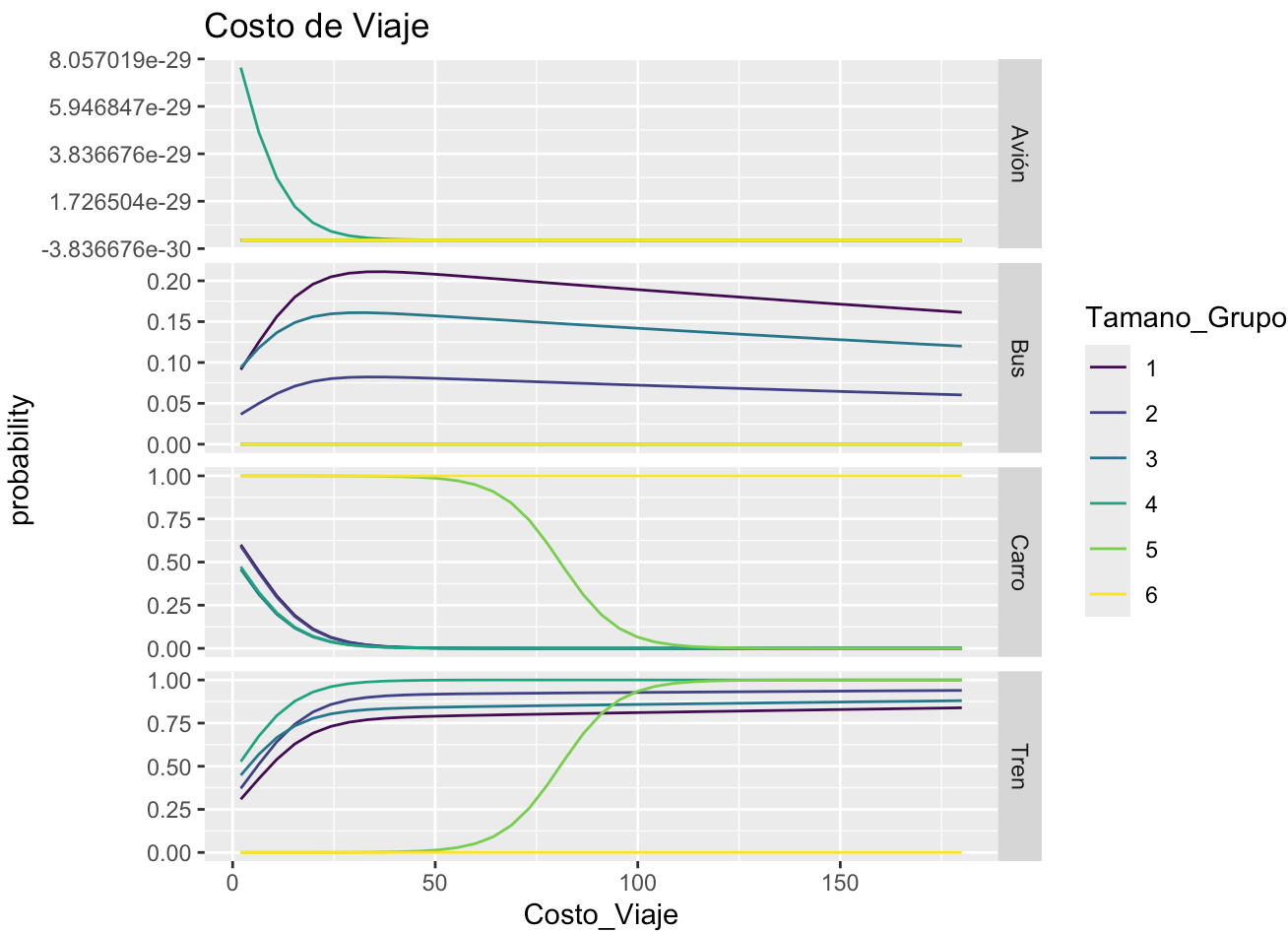
Avión	Bus	Carro	Tren
1.712114e-35	7.213883e-12	4.443178e-01	5.556822e-01

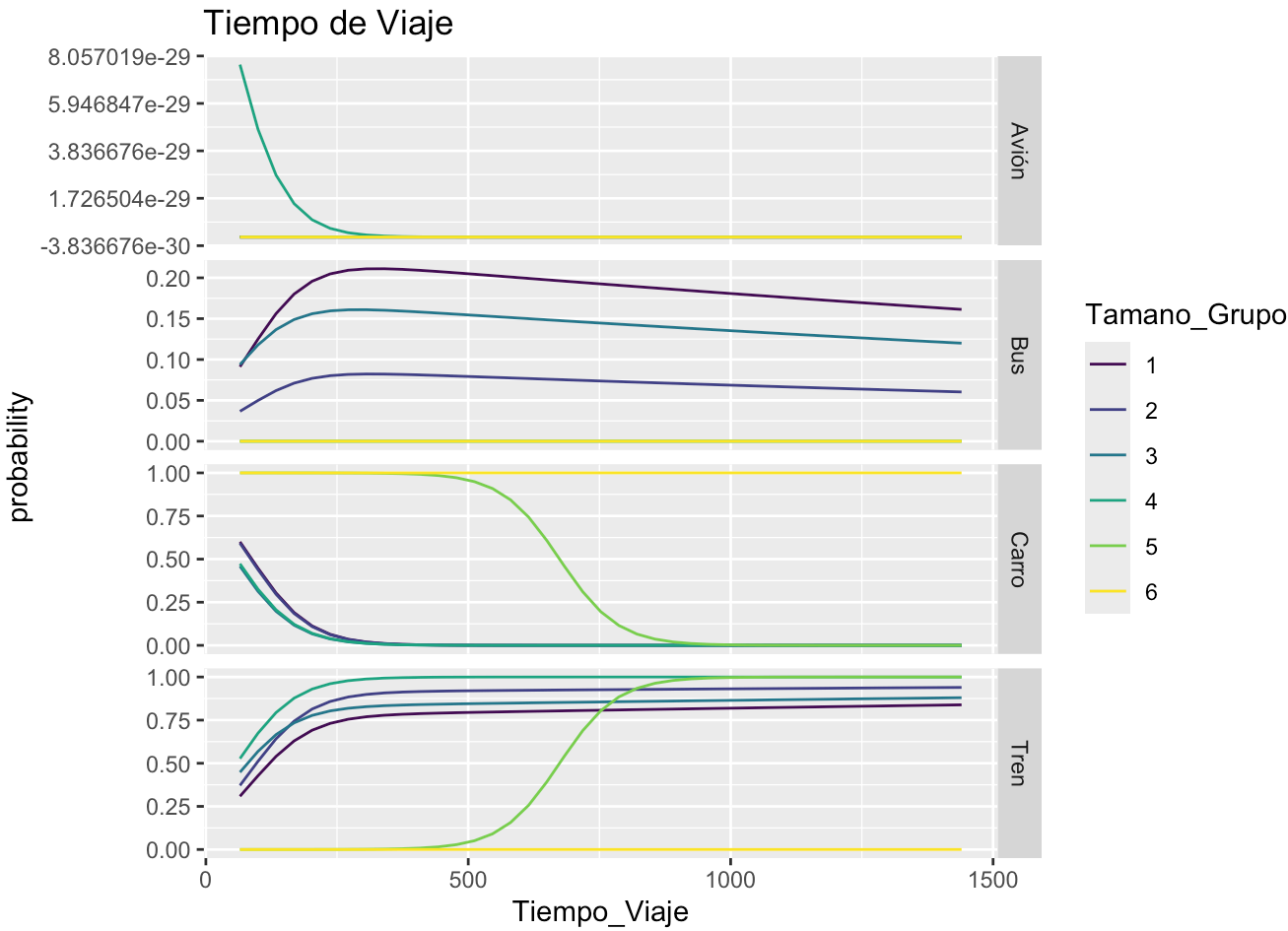
pp.sim\$Tamano_Grupo: 6

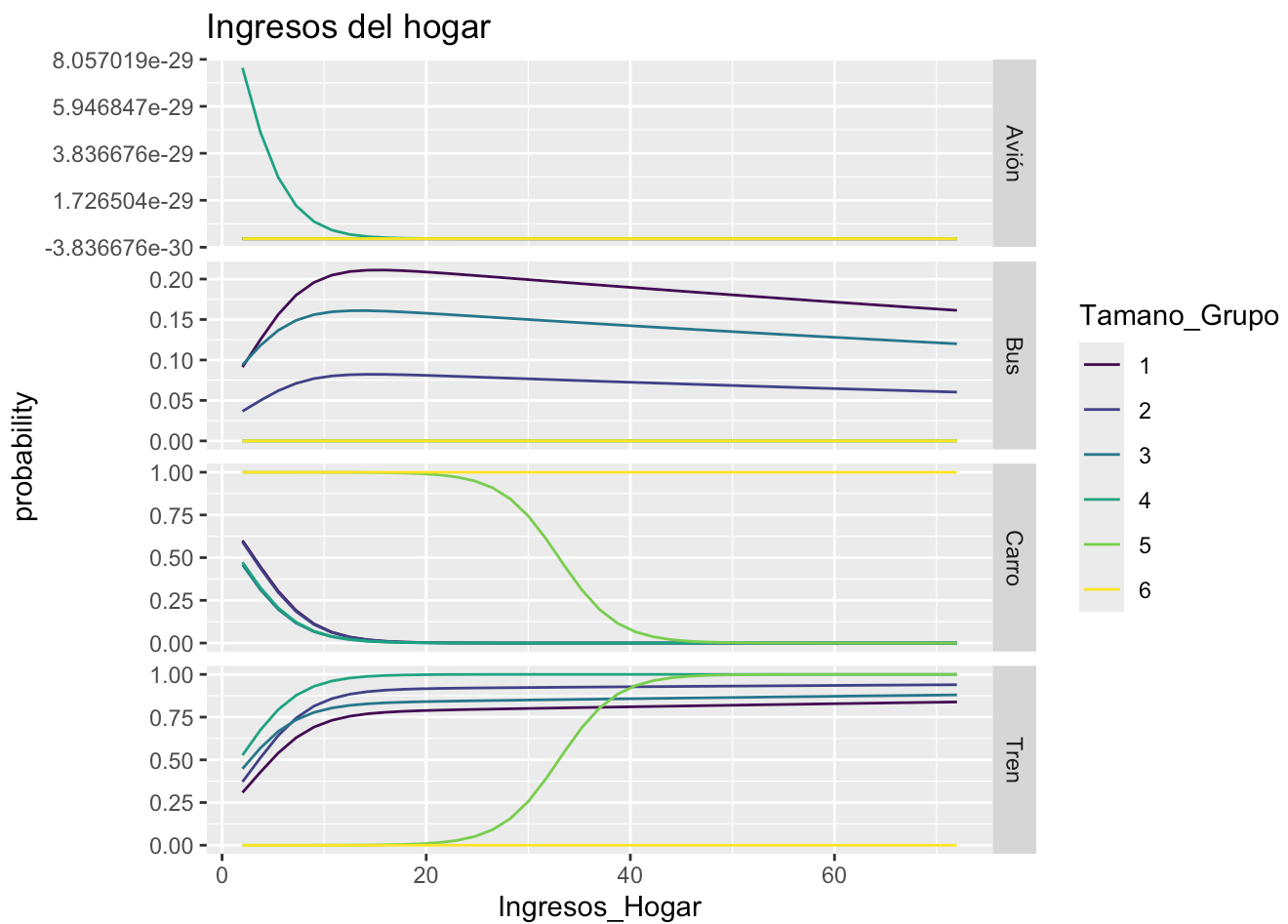
Avión	Bus	Carro	Tren
7.264613e-49	1.982124e-17	1.000000e+00	3.172226e-10

Poner hallazgos

Gráfico de probabilidades







Poner hallazgos

2.5. Diagnóstico

weights: 8 (3 variable)
 initial value 291.121816
 final value 283.758768
 converged

Pseudo R^2 : 0.6377087

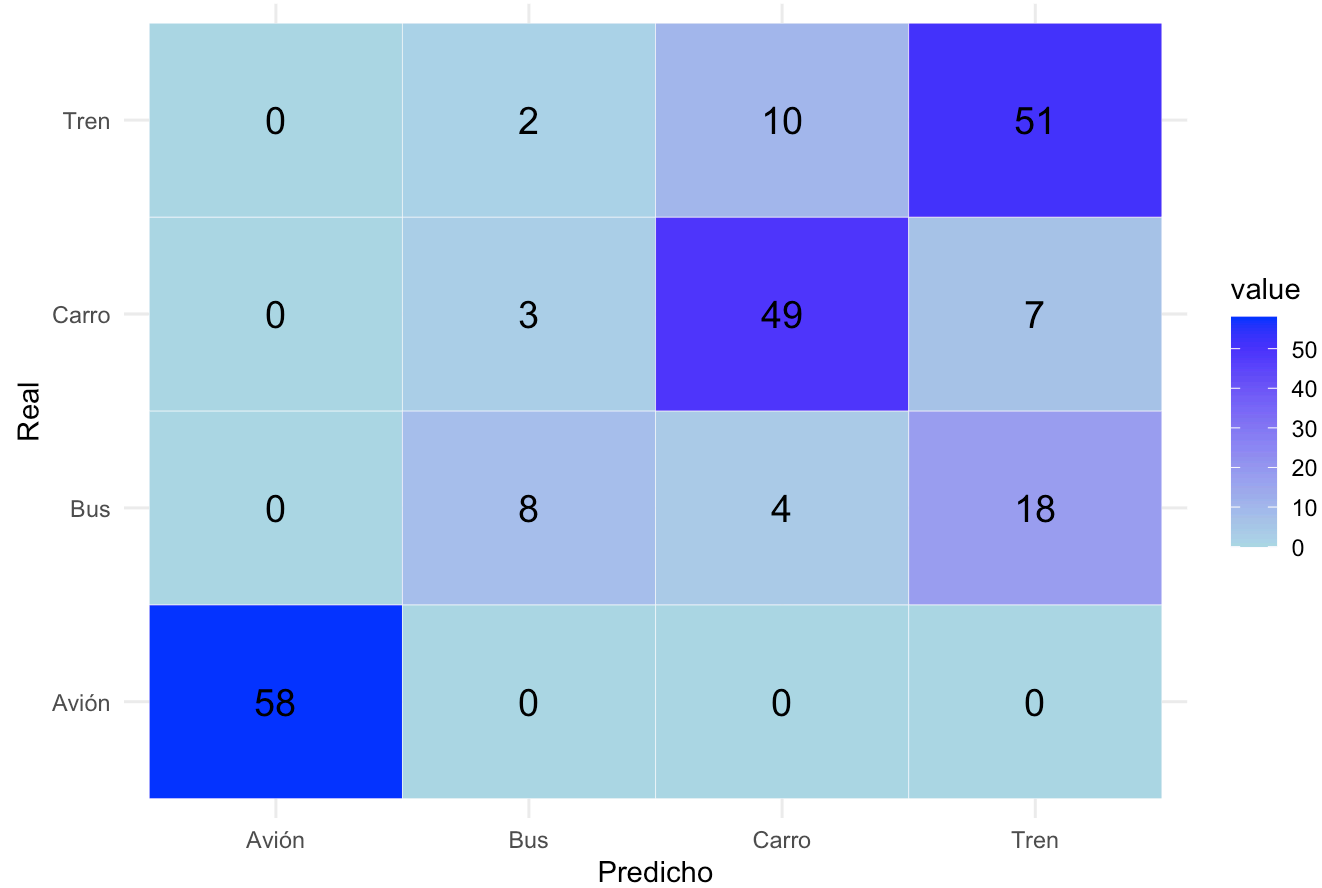
Log-likelihood: -102.8033

AIC: 259.6067

BIC: 349.9786

Matriz de confusión:

Matriz de Confusión - Modelo Multinomial



Reporte de clasificación:

Estadísticas generales

	x
Accuracy	0.7904762
Kappa	0.7098081
AccuracyLower	0.7291179
AccuracyUpper	0.8434390
AccuracyNull	0.3000000
AccuracyPValue	0.0000000
McnemarPValue	NaN

Estadísticas por clase

	Class: Avión	Class: Bus	Class: Carro	Class: Tren
Sensitivity	1.0000000	0.2666667	0.8305085	0.8095238
Specificity	1.0000000	0.9722222	0.9072848	0.8299320
Pos Pred Value	1.0000000	0.6153846	0.7777778	0.6710526

	Class: Avión	Class: Bus	Class: Carro	Class: Tren
Neg Pred Value	1.0000000	0.8883249	0.9319728	0.9104478
Precision	1.0000000	0.6153846	0.7777778	0.6710526
Recall	1.0000000	0.2666667	0.8305085	0.8095238
F1	1.0000000	0.3720930	0.8032787	0.7338129
Prevalence	0.2761905	0.1428571	0.2809524	0.3000000
Detection Rate	0.2761905	0.0380952	0.2333333	0.2428571
Detection Prevalence	0.2761905	0.0619048	0.3000000	0.3619048
Balanced Accuracy	1.0000000	0.6194444	0.8688966	0.8197279

Calcular Pseudo-R², log-likelihood, tasa de acierto (matriz de confusión) y analizar