Választási modellek 3

Prileszky István
Doktori Iskola
2018

http://www.sze.hu/~prile

Forrás: A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models Prepared For U.S. Department of Transportation Federal Transit Administration By Frank S. oppelman and Chandra Bhat with technical support from Vaneet Sethi, Sriram Subramanian, Vincent Bernardin and Jian Zhang January 31, 2006 Modified June
http://www.caee.utexas.edu/prof/bhat/COURSES/LM_Draft_060131Final-060630.pdf

A modell megfelelőségének vizsgálata

- A modell mennyire illeszkedik a minta (stated preference) elemeihez
  - informal judgment-based tests
  - goodness-of-fit measures
  - statistical tests
  - hit ratio

Variables	Zero Coefficients Model	Constants Only Model	Base Model
Travel Cost (1990 cents)			-0.0049 (-20.6)
Total Travel Time (minutes)			-0.0513 (-16.6)
Income (1,000's of 1990 DOLLARS)			
Drive Alone (Base)			0.0
Shared Ride 2			-0.0022 (-1.4)
Shared Ride 3+			0.0004 (0.1)
Transit			-0.0053 (-2.9)
Bike			-0.0128 (-2.4)
Walk			-0.0097 (-3.2)
Mode Constants			
Drive Alone (base)		0.0	0.0
Shared Ride 2		-2.137 (-44.1)	-2.178 (-20.8)
Shared Ride 3+		-3.303 (-40.6)	-3.725 (-21.0)
Transit		-1.950 (-38.5)	-0.6709 (-5.1)
Bike		-3.334 (-23.1)	-2.376 (-7.8)
Walk		-2.040 (-23.9)	-0.2068 (-1.1)
Log-likelihood at Zero		-7309.601	-7309.601
Log-likelihood at Constant			-4132.916
Log-likelihood at Convergence	-7309.601	-4132.916	-3626.186
Rho-Squared w.r.t. Zero	NA	0.4346	0.5039
Rho-Squared w.r.t. Constants	NA	NA	0.1226

Informal Tests

Logikai, szakmai összefüggések alapján mennyiben reálisak, valószerűek az eredmények

- Signs of Parameters (előjel)
  - Negatív olyan tényezőknél, amelyek csökkentik a választás valószínűségét (pl. utazási idő, költség)
  - Pozitív olyan tényezőknél, amelyek növelik a választás valószínűségét (pl. sebesség, komfort)
- Differences in Alternative Specific Variable Parameters across Alternatives
  - Egyes változók nagyobb szerepet játszanak bizonyos alternatíváknál, pl. nagyobb jövedelem jobban csökkenti a kényelmetlenebb alternatívák választását. Vizsgálhatjuk, hogy az eredmények igazolják-e ezt a feltevést. (Kivétel shared ride3+)

- The Ratio of Pairs of Parameters
  - Az együtthatók osztásával jellemző mutatószámokhoz juthatunk
    - Idő pénzértéke: „For example, in the Base Model reported in Table 5-2, the implied value of time is (-0.05134)/(-0.00492) or 10.4 cents per minute (the units are determined from the units of the time and cost variables used in estimation). This is equivalent to $6.26 per hour which is much lower that the average wage rate in the area at the time of the survey, approximately $21.20 per hour, suggesting that the estimated value of time may be too low.
    - out of vehicle time relative to in vehicle time
  - Winkler

Dr. Winkler

Tényező	**Együttható**	**Időérték***	**Pénzérték**	**„p” érték**
Átszállás (*ATSZ*)	-3,1000	11,7 perc	182,35 Ft	<0,01
Zsúfoltság (*ZS*)	-1,6500	6,23 perc	97,6 Ft	0,01
Gyaloglás (*GY*) (1 perc)	-0,4600	1,74 perc	27,06 Ft	<0,01
Várakozás (*VAR*) (1 perc)	-0,4230	1,6 perc	24,88 Ft	<0,01
Plusz díj (*FIZ*) (1 Ft)	-0,0170	0,06 perc	1 Ft	<0,01
Járművön töltött utazás (*JIDO*) (1 perc)	-0,2650	1 perc	15,59 Ft	<0,01
Konstans (*c*)	-0,0350	0,13 perc	2,03 Ft	***0,94***

* = járművön töltött utazási idő alapján

Az idő pénzértéke a járművön töltött utazás esetén óránként 935,4 Ft, várakozás esetén 1492,8 Ft, gyaloglás esetén 1623,6 Ft. Az átszállás időegyenértéke ~12 perc.

Overall Goodness-of-Fit Measures

- The rho-squared value (ρ^2) can be used to describe the overall goodness of fit of the model
- A loglikelihood függvény adatain alapul.
  - LL(0) represents the log-likelihood with zero coefficients (which results in equal likelihood of choosing each available alternative),LL(0) represents the log-likelihood for the constants only model, LL(βˆ) represents the log-likelihood for the estimated model and LL(*) = 0 is the log-likelihood for the perfect prediction model

Figure 5.2 Relationship between Different Log-likelihood Measures

Table 4-8 MNL Probabilities for Constants Only Model

Alternative	Utility		Exponent	Probability
	Expression	Value		
Drive Alone	0.0	0.0	1.0000	0.7314
Shared Ride	-1.60	-1.60	0.2019	0.1477
Transit	-1.80	-1.80	0.1653	0.1209
			1.3672	

Market shares model

$$\rho_0^2 = \frac{LL(\hat{\beta}) - LL(0)}{LL(*) - LL(0)}$$

Since the log-likelihood value for the perfect model is zero[23], the ρ_0^2 measure reduces to:

$$\rho_0^2 = 1 - \frac{LL(\hat{\beta})}{LL(0)}$$

Similarly, the rho-square with respect to the constants only model is:

$$\rho_c^2 = \frac{LL(\hat{\beta}) - LL(c)}{LL(*) - LL(c)}$$

$$= 1 - \frac{LL(\hat{\beta})}{LL(c)}$$

By definition, the values of both rho-squared measures lie between 0 and 1 (this is similar to the R2 measure for linear regression models). A value of zero implies that the model is no better than the reference model, whereas a value of one implies a perfect model; that is, every choice is predicted correctly.

- A problem with both rho-squared measures is that there are no guidelines for a "good" rho-squared value. (modellek összehasonlítása)
- Another problem with the rho-squared measures is that they improve no matter what variable is added to the model independent of its importance. A szabadságfokkal függ össze. One approach to this problem is to replace the rho-squared measure with an adjusted rho-square measure

$$\bar{\rho}_0^2 = \frac{\left[LL(\hat{\beta}) - K\right] - LL(0)}{LL(*) - LL(0)}$$

$$= 1 - \frac{LL(\hat{\beta}) - K}{LL(0)}$$

where K is the number of degrees of freedom (parameters) used in the model.

- The corresponding adjusted rho-squared for the constants only model

$$\bar{\rho}_C^2 = \frac{LL(\hat{\beta}) - K - (LL(C) - K_{MS})}{LL(*) - (LL(C) - K_{MS})} = 1 - \frac{LL(\hat{\beta}) - K}{LL(C) - K_{MS}} \qquad 5.12$$

where K_{ms} is the number of degrees of freedom (parameters) used in the constants only model.

Model:	Logit
Number of estimated parameters:	4
Number of observations:	6768
Number of individuals:	6768
Null log likelihood:	-6964.663
Init log likelihood:	-6964.663
Final log likelihood:	-5331.252
Likelihood ratio test:	3266.822
Rho-square:	0.235
Adjusted rho-square:	0.234
Final gradient norm:	+6.288e-004
Diagnostic:	Convergence reached...
Iterations:	4
Run time:	00:01
Variance-covariance:	from analytical hessian
Sample file:	swissmetro.dat

SM+Male

Model:	Logit
Number of estimated parameters:	5
Number of observations:	6372
Number of individuals:	6372
Null log likelihood:	-6529.613
Init log likelihood:	-6529.613
Final log likelihood:	-4874.995
Likelihood ratio test:	3309.235
Rho-square:	0.253
Adjusted rho-square:	0.253
Final gradient norm:	+2.520e-002
Diagnostic:	Convergence reached...
Iterations:	5
Run time:	00:00
Variance-covariance:	from analytical hessian
Sample file:	sm_tiszta9973.dat

Utility parameters

Name	Value	Std err	t-test	p-value	Robust Std err	Robust t-test	p-value
ASC_CAR	-0.338	0.0594	-5.68	0.00	0.0639	-5.29	0.00
ASC_SM	0.00	fixed					
ASC_TRAIN	-0.726	0.0590	-12.30	0.00	0.0937	-7.75	0.00
B_COST	-0.768	0.0675	-11.38	0.00	0.0810	-9.48	0.00
B_MALE	-0.0171	0.00275	-6.20	0.00	0.00338	-5.05	0.00
B_TIME	-1.41	0.0615	-22.84	0.00	0.121	-11.60	0.00

Statistical Tests

- Sztenderd hiba: $S = \frac{\sigma}{\sqrt{n}}$
- T-próba

The statistic used for testing the null hypothesis that a parameter $\hat{\beta}_k$ is equal to some hypothesized value, β_k^*, is the asymptotic t-statistic, which takes the following form:

$$t - statistic = \frac{\hat{\beta}_k - \beta_k^*}{S_k} \qquad 5.13$$

where	$\hat{\beta}_k$	is the estimate for the k^{th} parameter,
	β_k^*	is the hypothesized value for the k^{th} parameter and
	S_k	is the standard error of the estimate.

$\beta_k^* = 0$ null hipotézis

$$t = \frac{\beta_k^*}{S_k}$$

Confidence Level	Critical t-value (two-tailed test)
90%	1.645
95%	1.960
99%	2.576
99.5%	2.810
99.9%	3.290

The critical t values for different levels of confidence for samples sizes larger than 150 (which is the norm in mode choice analysis) are shown in Table 5-3.

One can conclude that a particular variable has no influence on choice (or equivalently that the true parameter associated with the variable is zero) can be rejected at the 90% level of confidence if the absolute value of the t-statistic is greater than 1.645 and at the 95% level of confidence if the t-statistic is greater than 1.96024.

p - value = 0,05

Table 5-3 Critical t-Values for Selected Confidence Levels and Large Samples

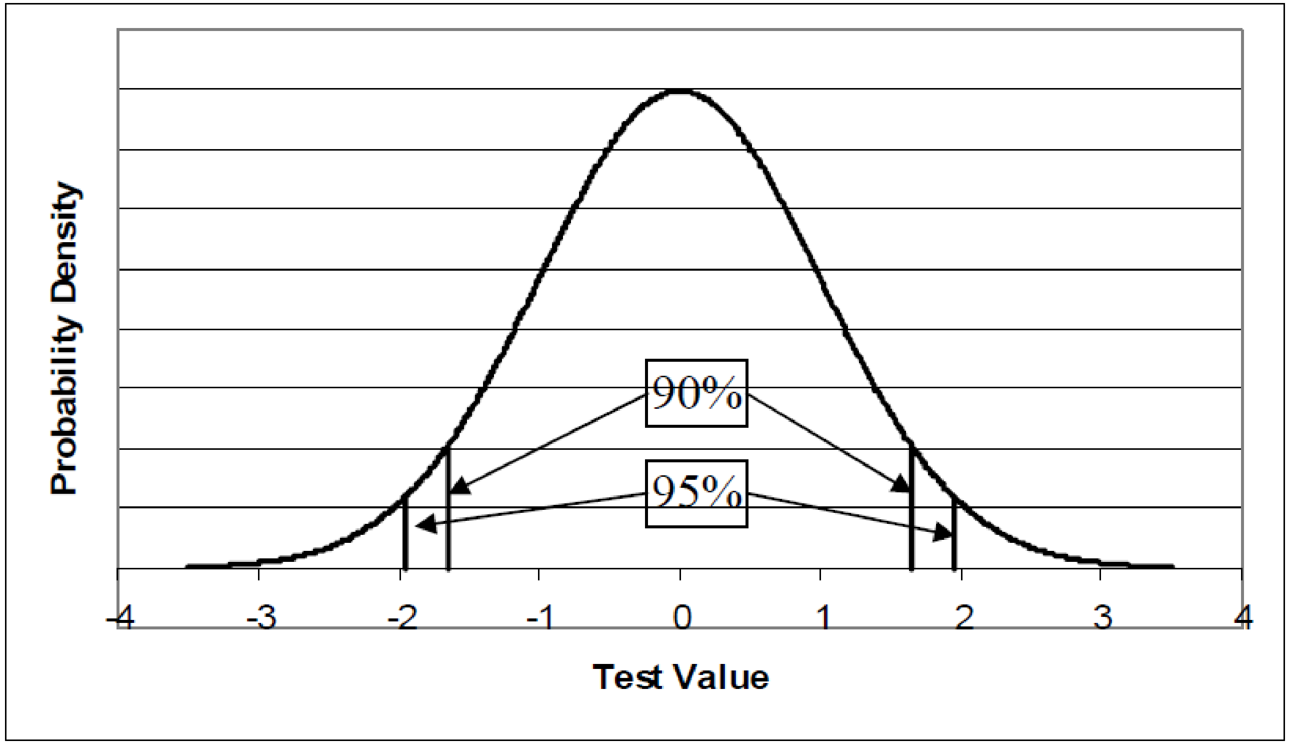

Figure 5.3 t-Distribution Showing 90% and 95% Confidence Intervals

- Korreláció –kovarencia
  - Két stisztikai adatsor kapcsolatát jellemzi a korrelációs együttható
  - Ha két attribútum adatsora között erős korreláció van akkor az egyik attribútum nem hordoz újabb információt a másikhoz képest. Érdemes vizsgálni, hogy nem egy harmadik ismérv alakulása befolyásolja-e ezek értékét.

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y},$$

  - Valószínűségi változóknál
    - Kovarencia

$$\mathrm{Cov}(X,Y) = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})$$

    - Korrelációs együtthato

$$r_{X,Y} = \mathrm{Cor}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{D_X^2 \cdot D_Y^2}$$

Correlation of coefficients

Coefficient1	Coefficient2	Covariance	Correlation	t-test	p-value		Rob. cov.	Rob. corr.	Rob. t-test	p-value	
B_COST	B_TIME	0.000550	0.187	2.79	0.01		0.00220	0.309	1.84	0.07	*
ASC_TRAIN	B_TIME	-0.00225	-0.722	5.56	0.00		-0.00760	-0.883	3.18	0.00	
ASC_TRAIN	B_COST	8.22e-006	0.00289	5.08	0.00		-0.000831	-0.147	3.34	0.00	
ASC_CAR	B_TIME	-0.00144	-0.585	12.57	0.00		-0.00482	-0.796	7.27	0.00	
ASC_CAR	B_COST	0.000485	0.216	15.52	0.00		2.86e-005	0.00722	10.40	0.00	
ASC_CAR	ASC_TRAIN	0.00138	0.580	11.85	0.00		0.00390	0.812	11.16	0.00	

- „Gyenge” változók sorsa – kivegyük a modellből, vagy ne?
- Közlekedésszakmai kérdés
  - Szakmai megítélés szerint van hatása?
  - A hasznossági függvény felhasználásakor a forgalom összetétele el fog térni az SP minta összetételétől?