Logistic regression instructions

Note that currently, the tool performs only binary logistic analysis with a binary outcome variable (i.e. linguistic variable with two categories).

	А	В	С	D	F	
1	ID	Variety	Genre	Subject	Outcome	Each line represents
2	1	B_BR	D_Press	A_other	A_other	a single observation
3	2	B_BR	D_Press	A_other	B_must	(i.e. occurrence of
4	3	B_BR	D_Press	A_other	A_other	feature in corpus).
5	4	B_BR	D_Press	A_other	A_other	
6	5	B_BR	D_Press	A_other	A_other	
7	6	B_BR	D_Press	A_other	A_other	
8	7	B_BR	D_Press	A_other	A_other	
9	8	B_BR	D_Press	C_you	A_other	One binary linguistic
10	9	B_BR	D_Press	A_other	A_other	variable ('Outcome')
11	10	B_BR	D_Press	A_other	B_must	and three predictor
12	11	B_BR	D_Press	A_other	A_other	variables.
13	12	B_BR	D_Press	A_other	B_must	
14	13	B_BR	D_Press	A_other	A_other	
15	14	B_BR	D_Press	A_other	A_other	
16	15	B_BR	D_Press	A_other	A_other	
17	16	B_BR	D_Press	A_other	A_other	

1) The tool expects input in the following format (only a small part of the dataset is shown).

2) One of the two options can be selected:

1. Select what you want to do. For help click $\underline{here.}$

Build a model 🔹							
Build a model							
Build a model stepwise							

The first option (default) selects a block entry of predictors (all relevant predictors at once); the other option selects step-wise predictor entry, which is useful especially for exploratory studies.

'BUILD A MODEL' OPTION

3a) Follow the instructions on the screen and enter all relevant pieces of information. Press 'Build model'.

	lel 🔻														
2. Paste data in tł	ne text area.														
1938 A_AM B_F 1939 A_AM B_F 1940 A_AM B_F 1941 A_AM B_F 1942 A_AM B_F 1942 A_AM B_F 1943 A_AM B_F 1944 A_AM B_F 1946 A_AM B_F 1946 A_AM B_F 1947 A_AM B_F 1949 A_AM B_F 1949 A_AM B_F 1950 A_AM B_F 1951 A_AM B_F	Iction A_other Iction A_other Iction A_other Iction B_I Iction A_other Iction A_other Iction A_other Iction A_other Iction A_other Iction B_I Iction C_you Iction C_you Iction A_other	A_other A_other A_other A_other A_other A_other A_other A_other A_other A_other A_other A_other A_other A_other A_other													
3. Type in the exa	act name of the ou	utcome varia	able. <mark>Outcom</mark>	ie											
4. Type in the exa	act name(s) of the	predictor(s) [use ; as s	eparato	or]. Variety;	Genre									
5. Decide if you w	ant to include pre	edictor intera	actions.												
• Yes, include all	Yes, include so	ome 🔍 No													
	- res, merade se														
 Type in the exa 	act names of the p	oredictors wi	th interactio	ns Luse	e ; as separa	tor]. [V	/ariety;	Ger	nre						
Build model Cla	ar.					T ł	ne in	te	ract	ions					
Build Hodel Cle	:41				Build model Clear The interactions										
setting is optional.															
						L I	f int	er	este	d only					
			-	-	-	I I ma	f int ain e	er ff	este ects	d only , answe					
			-		-	I i ma	f int ain e No' i	er ff n	este ects Step	d only , answe 5.					
			-			t I ma 1'	f int ain e No' i	er ff n	este ects Step	d only , answe 5.					
he following is a	an example of th	ne output:				t I sm 1'	f int ain e No' i	er ff n	este ects Step	d only , answe 5.					
he following is a	in example of th	ne output:				I f ma 'h Su	f int ain e No' i ummar	er ff n y	este ects Step stat	d only , answe 5. s for t					
he following is a ut	in example of th	ne output:				It ma 'l Su wł	f int ain e No' i ummar	er ff n Y mo	este ects Step stat del.	d only , answe 5. s for t					
he following is a ut Il model statistics: Lik	an example of th	ne output: : 145.56 (p < 0	.0001) -> SIGN	IFICANT	; C-index: 0.66	I f ma 'I Su wł	f int ain e No' i ummar nole :	er ff n y mo	este ects Step stat del.	d only , answe 5. s for t					
he following is a ut Il model statistics: Lik VIC: 2515.61	an example of th elihood ratio test (LL):	ne output: : 145.56 (p < 0 Signifi	.0001) -> SIGN	IFICANT	; C-index: 0.66 E.ffe	If ma 'n Su Wh -> NOT	f int ain e No' i ummar nole : ACCEPTA	er ff n y mo	este ects Step stat del.	d only , answe 5. s for t					
he following is a ut Il model statistics: Lik IIC: 2515.61 cients:	an example of th	ne output: : 145.56 (p < 0 Signifi	.0001) -> SIGN LCance	IFICANT	;C-index: 0.66 Effe	Ii ma 'n Su wh -> NOT ct s:	f int ain e No' i ummar nole : ACCEPTA ize	er ff n Y mo	este ects Step stat del.	d only , answe 5. s for t					
he following is a ut Il model statistics: Lik IC: 2515.61 cients:	elihood ratio test (LL):	ne output: : 145.56 (p < 0 Signifi Standard Error	.0001) -> SIGN LCance	IFICANT	; C-index: 0.66 Effe	Si If ma 'l Si Wł -> NOT ct s:	f int ain e No' i ummar nole ACCEPTA ize	er ff n y mo	este ects Step stat del.	d only , answe 5. s for t					
he following is a ut Il model statistics: Lik IC: 2515.61 cients:	elihood ratio test (LL): Estimate (log odds)	ne output: : 145.56 (p < 0 Signifi Standard Error	.0001) -> SIGN i Cance Z value (Wald)	IFICANT;	; C-index: 0.66 Effe =stimate (odds	Si If ma 'I Si WI -> NOT ct s: -> NOT -> NOT -> NOT	f int ain e No' i ummar nole : ACCEPTA ize	er ff n y mo	este ects Step stat del. ; Nagelk	d only , answe 5. s for t cerke R ² :					
he following is a ut II model statistics: Lik IIC: 2515.61 cients: rcept) tvB BR	elihood ratio test (LL): Estimate (log odds) 0.542 -0.930	ne output: 145.56 (p < 0 Signifi Standard Error 0.200 0.290	.0001) -> SIGN i Cance Z value (Wald) 2.714 -3.210	IFICANT; p-value 0.007 0.001	; C-index: 0.66 Effe Estimate (odds 1.720).395	I f ma 'Y St wh -> NOT ct s: [ower 1.168 0.222	f int ain e No' i ummar nole : ACCEPTA ize 95% Cf upper 2.561 0.693	er ff n y mo	este ects Step stat del. ; Nagelk	d only , answe 5. s for t cerke R ² :					
he following is a ut II model statistics: Lik IIC: 2515.61 cients: rcept) tyB_BR BR BR_Fiction	elihood ratio test (LL): Estimate (log odds) 0.542 -0.930 -1.318	ne output: 145.56 (p < 0 Signifi Standard Error 0.200 0.290 0.242	.0001) -> SIGN i Cance Z value (Wald) 2.714 -3.210 -5.441	P-value 0.007 0.001	; C-index: 0.66 Effe Estimate (odds 1.720).395).268	I f ma 'Y St wh -> NOT ct s: 0.222 0.165	f int ain e No' i ummar nole ACCEPTA ize 95% Cf upper 2.561 0.693 0.693	er ff n y mo	este ects Step stat del. ; Nagelk Eff	d only , answe 5. s for t cerke R ² :					
he following is a ut II model statistics: Lik IIC: 2515.61 cients: rcept) tyB_BR BR_Fiction ac_General	elihood ratio test (LL): Estimate (log odds) 0.542 -0.930 -1.318 -0.450	ne output: 145.56 (p < 0 Signifi Standard Error 0.200 0.290 0.242 0.228	.0001) -> SIGN Cance Z value (Wald) 2.714 -3.210 -5.441 -1.974	P-value 0.007 0.001 0.000 0.048	; C-index: 0.66 Effe Estimate (odds 1.720).395).268).638	I 1 ma 'Y St wh -> NOT ct s: 0.222 0.165 0.406	f int ain e No' i ammar nole ACCEPTA ize 95% CI upper 2.561 0.693 0.693	er ff n y mo	este ects Step stat del. ; Nagelk Eff ind pre	d only , answe 5. s for t cerke R ² :					
ne following is a it II model statistics: Lik IC: 2515.61 cients: rcept) tyB_BR B_Fiction aC_General aD_Press	elihood ratio test (LL): Estimate (log odds) 0.542 -0.930 -1.318 -0.450 -1.191	ne output: 145.56 (p < 0 Signifi Standard Error 0.200 0.242 0.228 0.268	.0001) -> SIGN Cance Z value (Wald) 2.714 -3.210 -5.441 -1.974 -4.445	P-value 0.007 0.001 0.0048 0.000	; C-index: 0.66 Effe Estimate (odds 1.720).395).268).638).304	I 1 ma 'N St wh -> NOT ct s: 0.222 0.165 0.222 0.165	f int ain e No' i ammar nole ACCEPTA ize 95% CI upper 2.561 0.693 0.693 0.693 0.511	y mo	este ects Step stat del. ; Nagelk Eff ind pre	d only , answe 5. s for t cerke R ² :					
he following is a ut II model statistics: Lik IIC: 2515.61 cients: rcept) tyB_BR BR_Fiction aC_General aD_Press ctB_I	elihood ratio test (LL): Estimate (log odds) 0.542 -0.930 -1.318 -0.450 -1.191 -1.084 -0.97	<pre>he output: 145.56 (p < 0 Signifi Standard Error 0.200 0.290 0.242 0.268 0.174 0.150</pre>	.0001) -> SIGN Cance Z value (Wald) 2.714 -3.210 -5.441 -1.974 -4.445 -6.232 5.704	P-value 0.007 0.001 0.000 0.048 0.000 0.048	; C-index: 0.66 Effe Estimate (odds 1.720).395).268).638).304).338	Si I 1 ma ' N Si Si wh -> NOT ct s: 95% Cl lower 1.165 0.222 0.165 0.406 0.179 0.232 0.221 0.222 0.165 0.406 0.179 0.221 0.221 0.222 0.22 0.222 0.22 0.222 0.22	f int ain e No' i ammar nole ACCEPTA ize 95% CI upper 2.561 0.693 0.693 0.693 0.693 0.511 0.428	y mo	este ects Step stat del. ; Nagelk Eff ind pre	d only , answe 5. s for t cerke R ² :					
he following is a ut II model statistics: Lik IIC: 2515.61 cients: rcept) tyB_BR BR_Fiction aC_General aD_Press ictB_I ictC_you tyB_B E:GeneraB_Eiction	elihood ratio test (LL): Estimate (log odds) 0.542 -0.930 -1.318 -0.450 -1.191 -1.084 -0.917 -0.1482	ne output: 145.56 (p < 0 Signifi Standard Error 0.200 0.242 0.228 0.268 0.174 0.158 0.340	.0001) -> SIGN Cance Z value (Wald) 2.714 -3.210 -5.441 -1.974 -4.445 -6.232 -5.794 4 353	D-value 0.007 0.001 0.000 0.048 0.000 0.000 0.000	; C-index: 0.66 Effe Estimate (odds 1.720).395).268).638).304).338).400	I 1 ma 'N Sl wh -> NOT ct s: 0.222 0.165 0.222 0.165 0.239 0.239 0.239	f int ain e No' i ammar nole ACCEPTA ize 95% CI upper 2.561 0.693 0.693 0.693 0.693 0.511 0.472 0.542 8.620	y mo	este ects Step stat del. ; Nagelk Eff ind pre	d only , answe 5. s for t cerke R ² :					
he following is a ut II model statistics: Lik IIC: 2515.61 (cients: rcept) tyB_BR eB_Fiction eC_General eD_Press setB_I setC_you tyB_BR:GenreB_Fiction tyB_BR:GenreC_Gener	Estimate (log odds) 0.542 -0.930 -1.318 -0.450 -1.191 -1.084 -0.917 n 1.482 -alo .952	ne output: 145.56 (p < 0 Signifi Standard Error 0.200 0.242 0.228 0.268 0.174 0.158 0.340 0.324	.0001) -> SIGN Cance 2 value (Wald) 2.714 -3.210 -5.441 -1.974 -4.445 -6.232 -5.794 4.353 2.941	P-value 0.007 0.001 0.048 0.000 0.048 0.000 0.000 0.000 0.000	; C-index: 0.66 Effe Estimate (odds 1.720).395).268).304).304).338).400 1.400 2.592	<pre>I 1 1 ma</pre>	f int ain e No' i ammar nole ACCEPTA ize 95% CI upper 2.561 0.693 0.693 0.693 0.693 0.693 0.691 0.693 0.511 0.472 0.542 8.620	y mo BLE	este ects Step stat del. ; Nagelk Eff ind pre	d only , answe 5. s for t cerke R ² :					
he following is a ut III model statistics: Lik IIC: 2515.61 icients: rcept) tyB_BR eS_Fiction eC_General eD_Press ictB_I ictC_you tyB_BR:GenreB_Fiction tyB_BR:GenreC_Gener tyB_BR:GenreD_Press	elihood ratio test (LL): Estimate (log odds) 0.542 -0.930 -1.318 -0.450 -1.191 -1.084 -0.917 n 1.482 -a) 0.952 1.490	145.56 (p < 0 Signifi Standard Error 0.200 0.242 0.228 0.268 0.174 0.340 0.324 0.324 0.379	.0001) -> SIGN Cance 2.714 -3.210 -5.441 -1.974 -4.445 -6.232 -5.794 4.353 2.941 3.927	D-value 0.007 0.001 0.000 0.048 0.000 0.048 0.000 0.000 0.000 0.000 0.000	; C-index: 0.66 Effe Estimate (odds 1.720).395).268).304).304).338).400 1.400 2.592 1.438	<pre>I i i i i i i i i i i i i i i i i i i i</pre>	f int ain e No' i ammar nole ACCEPTA ize 95% CI upper 2.561 0.693 0.693 0.693 0.693 0.693 0.691 0.693 0.511 0.472 0.542 8.620 8.620	y moo	este ects Step stat del. ; Nagelk Eff ind pre	d only , answe 5. s for t cerke R ² :					
he following is a ut all model statistics: Lik AIC: 2515.61 icients: rcept) tyB_BR eC_General eD_Press ectB_I ectC_you tyB_BR:GenreB_Fiction tyB_BR:GenreD_Press ession equation (predic	elihood ratio test (LL): Estimate (log odds) 0.542 -0.930 -1.318 -0.450 -1.191 -1.084 -0.917 n 1.482 -al 0.952 1.490 ted probabilities):	ne output: 145.56 (p < 0 Signifi Standard Error 0.200 0.242 0.228 0.242 0.228 0.174 0.158 0.340 0.324 0.379	.0001) -> SIGN Cance Z value (Wald) 2.714 -3.210 5.441 -1.974 -4.445 -6.232 -5.794 4.353 2.941 3.927	IFICANT, 0.007 0.001 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000	; C-index: 0.66 Effe 5stimate (odds 1.720).395).268).304).304).338).400 1.400 2.592 1.438	<pre>If it is it it</pre>	f int ain e No' i ummar hole ACCEPTA ize 95% CI upper 2.561 0.693 0.428 0.993 0.511 0.511 0.511 0.512 8.620 4.915 9.384	y mon BLE	este ects Step stat del. ; Nagelk Eff ind pre	d only , answe 5. s for t cerke R ² :					
he following is a ut all model statistics: Lik AIC: 2515.61 icients: rcept) tyB_BR eB_Fiction eC_General eD_Press actB_I actD_you tyB_BR:GenreB_Fiction tyB_BR:GenreD_Press ission equation (predic	elihood ratio test (LL): Estimate (log odds) 0.542 -0.930 -1.318 -0.450 -1.191 -1.084 -0.917 n 1.482 -0.952 1.490 ted probabilities):	ne output: 145.56 (p < 0 Signifi Standard Error 0.200 0.242 0.228 0.242 0.228 0.174 0.158 0.174 0.158 0.340 0.324 0.379	.0001) -> SIGN Cance Z value (Wald) 2.714 -3.210 -5.441 -1.974 -4.445 -6.232 -5.794 4.353 2.941 3.927	IFICANT, 0-value 0.007 0.001 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000	; C-index: 0.66 Effe 5stimate (odds 1.720).395).268).304).338).400 1.400 2.592 1.438	<pre>If it it</pre>	f int ain e No' i ammar hole ACCEPTA ize 95% CI upper 2.561 0.693 0.511 0.542 0.593 0.511 0.542 8.620 4.915 9.384	y moo	Eff ind pre	d only , answe 5. s for t cerke R ² : fect of lividua dictor					

Equation for calculating predicted probabilities.

'BUILD A MODEL STEPWISE' OPTION

3b) Follow the instructions on the screen and enter all relevant pieces of information. Press 'Build model'.

 2. Paste data in the text area. ID Variety Genre Subject Outcome B_BR D_Press A_other A_other 3. Type in the exact name of the outcome variab 	le. <mark>Outcome</mark>
4. Decide on the method of step-wise model buil	'Forward' means adding predictors
●Forward ◎Backward ◎Hybrid	one by one to a null model (with no predictors); 'Backward means deleting predictors one by one
Build model Clear	<pre>from a full model (with all predictors); 'Hybrid' combines these two procedures allowing adding or deleting predictors at each stage.</pre>

4a) The following is an example of the output:

Output

Summary of the forward step-wise procedure:

Step	Df	Deviance	Residual df	Residual deviance	AIC
1	NA	NA	1950	2641.168	2643.168
2+Subject	-2	87.446051	1948	2553.721	2559.721
3+Genre	-3	32.571310	1945	2521.150	2533.150
4+Variety	-1	2.554401	1944	2518.596	2532.596

Overview of all the steps taken to arrive at the 'best model'. '+' indicates adding one variable at a time until AIC stops decreasing. '-' indicates deleting one variable at a time until AIC stops decreasing. The smaller the AIC value, the better the model.

Best model - overall statistics: Likelihood ratio test (LL): 122.57 (p < 0.0001) -> SIGNIFICANT; C-index: 0.64 -> NOT ACCEPTABLE; Nagelkerke R²: 0.08; AIC: 2532.6

Coefficients:

Summary stats for the whole model.

	Estimate (log odds)	Standard Error	Z value (Wald)	p-value	Estimate (odds)	95% CI	95% CI
	Lotinace (log oddo)				Loannace (oddo)	lower	upper
(Intercept)	0.030	0.148	0.202	0.840	1.030	0.770	1.379
SubjectB_I	-1.060	0.173	-6.120	0.000	0.346	0.245	0.483
SubjectC_you	-0.905	0.157	-5.744	0.000	0.405	0.295	0.548
GenreB_Fiction	-0.590	0.169	-3.485	0.000	0.554	0.398	0.772
GenreC_General	-0.015	0.160	-0.097	0.923	0.985	0.720	1.347
GenreD_Press	-0.467	0.186	-2.504	0.012	0.627	0.435	0.903
VarietyB_BR	0.153	0.096	1.597	0.110	1.165	0.966	1.405

Effect of individual predictors (see above).

[Notes]