



LIMO EEG: what are robust statistics? Application to a categorical design

D2.A1 & D2.A3. - 2.00 to 3.30

Cyril Pernet, PhD Centre for Clinical Brain Sciences, The university of Edinburgh, UK

SCIENTIFIC DATA

OPEN SUBJECT CATEGORIES » Electroencephalography A multi-subject, multi-modal human neuroimaging dataset

-EEG

Daniel G. Wakeman^{1,2} & Richard N. Henson²

• Scientific Data 2, Article number: 150001 (2015)

doi:10.1038/sdata.2015.1

» Brain imaging

The Data

- 3 types of stimuli: Famous faces, Non-famous faces, Scrambled faces
- 3 levels of repetition: 1st time, 2nd time (right after), 3rd time (delayed)
- →Priming experiment with a possible interaction with the type of stimuli.

We need the conditions computed per subject (1st level) and then do the repeated measure ANOVA to test main effects and interactions.

What are we going to do?

• 1 – Replicate Henson et al. – faces vs. scrambled



- 2 learn about robust statistics
- 3 see how to extend to a 3 (category) by 3 (repetition) design

Let's get started

- Open Matlab
- Move to the data
- \geq 18 subjects
- type 'eeglab'
- Load/Make the 'study'
- \succ file \rightarrow load existing study



ds117_sub002_p05_famous_epochs.fdt ds117_sub002_p05_famous_epochs.set ds117_sub002_p05_scrambled_epochs.fdt ds117_sub002_p05_scrambled_epochs.set ds117_sub002_p05_unfamiliar_epochs.fdt ds117_sub002_p05_unfamiliar_epochs.set

~

📣 MATLAB R2013a

HOME

PLOTS

APPS

🛃 EE	GLAB v14.x (dev)		×
File	Edit Tools Plot	Study Datasets Help	
	STUDV set' 1	Edit study info	
	OTODT Set. (Select/Edit study design(s)	
		Precompute channel measures	
	Study filena	Plot channel measures	
	Study task r	LInear MOdeling EEG Data (LIMO/Channels)	
	Nb of subjed	Precompute component measures	
	Nh of condit	PCA clustering (original)	
	ND OI CONGIO	Edit/plot clusters	
	Nb of sessio	LInear MOdeling EEG Data (LIMO/Components) >	
	Nb of groups	a 1 per subject	
	Epoch consis	stency yes	
	Channels per	frame 60,65,66,67,68,69,70	
	Channel loca	ations yes	
	Clusters	21	
	Status	Pre-clustered	
	Total size	(Mb) 228.1	

warning. Stop: moved since last saved, tryin
pop_loadset(): loading file E:\My_Documents'
Warning: STUDY moved since last saved, tryin
pop_loadset(): loading file E:\My_Documents'
Warning: STUDY moved since last saved, tryin
pop_loadset(): loading file E:\My_Documents'
Warning: STUDY moved since last saved, tryin
pop_loadset(): loading file E:\My_Documents'
STUDY Warning: the trial information collect
Done.
Re-saving study file
Done.
fx >>

X 🖪 🔚 🔏 🖻 🗇 🗇 🔁 🕐 Search Documentation X _ 0 Create a new STUDY set -- pop_study() Edit STUDY set information - remember to save changes STUDY set name: - P wakeman henson study () STUDY set task name: ScrambledVsNormalFace STUDY set notes: ct> dataset filename subject condition browse session group Select by r.v. E:\My Documents\MATLAB\F 600,1] sub002 1 familiar 1 Comp.: 1 2. Clear 600 11 and repetition/acaloub002/da E:\My Documents\MATLAB\F 2 sub002 1 unfamiliar 1 Comp.: 1 2 .. Clear and repetition/ang/aub002/da () E:\My Documents\MATLAB\F 3 sub002 1 scrambled 1 Comp.: 1 2. Clear Date Modified and repetition/angloub/002/da E:\My Documents\MATLAB\F 5/10/2016 23:23 4 Comp.: 1 2. sub003 1 familiar 1 Clear and repetition/ang/outb003/da 5/10/2016 20:00 E:\My_Documents\MATLAB\F 5 1 sub003 unfamiliar 1 Comp.: 1 2. Clear 5/10/2016 17:56 and repetition/anglaub002/da 5/10/2016 18:00 E:\My Documents\MATLAB\F 6 1 sub003 scrambled 1 Comp.: 1 2. Clear 5/10/2016 18:05 and repetition/acaloub/002/da 5/10/2016 18:09 E:\My Documents\MATLAB\F 1 sub004 familiar 1 Comp.: 1 2 .. Clear 5/10/2016 18:14 8 E:\My Documents\MATLAB\F 1 15/10/2016 18:19 sub004 unfamiliar 1 Comp.: 1 2. Clear and repetition/angloub/00/1/da 15/10/2016 18:23 E:\My Documents\MATLAB\F Comp.: 1 2 sub004 1 scrambled 1 Clear 15/10/2016 20:00 and repetition/ang/outb004/da 15/10/2016 19:58 E:\My_Documents\MATLAB\F 10 sub005 familiar 1 Comp.: 1 2. Clear 15/10/2016 20:03 and repetition/acaloub005/da Important note: Removed datasets will not be saved before being deleted from EEGLAB memory 5/10/2016 20:07 5/10/2016 20:17 Page 1 < > 5/10/2016 20:21 5/10/2016 20:26 5/10/2016 20:30 Dataset info (condition, group, ...) differs from study info. [set] = Overwrite dataset info for each dataset on disk. 5/10/2016 20:35 7/10/2016 21:23 Delete cluster information (to allow loading new datasets, set new components for clustering, etc.) Ok Help Cancel 111

MATLAB R2013	3a										X
HOME	PLOTS	APPS								🖥 🗇 🖻 🔁 Seard	ch Documentation 🛛 🔎
EEGLAB v14	4.x (dev)	A			Select a	and compute component n	neasures for later clustering pop_preco	omp()			
File Edit T	Fools Plot	Study Datasets H	lelp	ъ <mark>1</mark>							
0.71		Edit study info		F	Pre-	compute channel m	neasures for STUDY 'wakema	n_henson_study'			
siu	JDY set: V	Select/Edit stud	y design(s)								-
		Precompute cha	nnel measures			Spherical interpolati	on of missing channels (perform	ned after optional ICA removal below	/)		•
Stud	dy filena	Plot channel me	asures	1		Remove ICA artifac	tual components pre-tagged in a	each dataset		Val	10
Stud	dy task n	LInear MOdeling	g EEG Data (LIMO/Channels)							val	
Nb o	of subjec	Precompute cor	nponent measures			Remove artifactual	CA cluster or clusters (hold shif	t key) Parentcluster 1		<1x <1x	54 struct>
Nbo	of condit	Edit/plot cluster	onginal) .s						-	10	
Nb o	of sessic	Linear MOdeling	EEG Data (LIMO/Components)	•						16	
Nbo	of groups		1 per subject		List	of measures to pre	compute			8	300.0.9600.11
Epoc	ch consis	tency	Ves	1		EDDa	Peopline (Imin movil in mo)			10.6	<u>600 0 7600 11</u>
Char	nnels ner	frame	60.65.66.67.68.69.70			ERFS				Current Fol	dor.
Char	nnol loga	tiona	vo.			Power spectrum	Spectopo parameters	'specmode', 'fft', 'logtrials', 'off'	Test	Current ron	Date Modified
cliai		CIONS	yes	1		ERP-image	ERP-image parameters	'nlines' 10 'smoothing' 10	Test		14/10/2016 23:45
CIUS	sters		1	1		5000		innice, ie, enteetinig, ie			14/10/2016 23:45
Stat	tus		Pre-clustered			ERSPS	Time/freg. parameters	'cycles' [3.0.8] 'nfregs' 100 'ntin	Test	ous_epochs.fdt	14/10/2016 23:45
Tota	al size (1	Mb)	228	1		ITCs	1.1-			ous_epochs.set	14/10/2016 23:45
						L				mbled_epochs.tdt	14/10/2016 23:45
										amiliar epochs.fdt	14/10/2016 23:45
pop_lo	adset (): loading	file E:\My_Doc	uments\]	<u>ک</u> 0۱	verwrite files on disk				amiliar_epochs.set	14/10/2016 23:45
pop_lo	adset (): loading	file E:\My_Doc	uments\							
pop_lo	adset (): loading	file E:\My_Doc	uments\							
pop lo	adset (): loading	file E:\My Doc	uments\]	H	leip		Cancel	Ok		
STUDY	Warnin	a: the tri	al information	collecte	VI LIV	m acabeen m	as changedy use prov			J	
Done		<u> </u>									
Done.											

We now extract all the single trials

-- no interpolation of missing channels ; LIMO EEG handles missing data

-- data were processed with epochs -200/800



The new STUDY allows all sorts of designs By default, you can model each and any condition / covariate Here, we pick the categorical variable 'condition'

EE	GLAB v14.x (dev)	c-a-d	1.0				<pre>B\Face_repetition\e</pre>		🛿 LInear MOdeling of EEG data pop_limo()
ile	Edit Tools Pla	ot Stu	ıdy	Datasets Help		ъ	B\Face_repetition\e		
			Ed	it study info		_	B\Face_repetition\e		Linear MOdeling of EEG data of design1
	01001300	· '	Se	ect/Edit study design(s)			B\Face repetition\e		(Use STUDY design interface to switch to a different design)
			Pre	compute channel measures			B\Face repetition\e		
	Study IIIe	r	LIn	ear MOdeling EEG Data (LIMO/Channels)	Estir	ima	ate Model Parameters (Channels) n\e		
	Nh of subi		Pre	ecompute component measures	Line	ear	Model Results (Channels) n \ e		Input data to use for the GLM ERP
	Nb of cond	it	PC.	A clustering (original)			B\Face repetition\e		Options 'timelim' [-50 650]
	Nb of sess	ic	Ed	It/plot clusters ear MOdeling EEG Data (LIMO/Components)			B\Face_repetition\e		Optimization method WLS
	Nb of grou	ps		1 per subject			B\Face_repetition\e		Frase previous model for this design and measure
	Epoch cons	ister	псу	yes			<pre>B\Face_repetition\e</pre>		
	Channels p	er fi	ram	e 60,65,66,67,68,69,70			<pre>B\Face_repetition\e</pre>		
	Channel lo	catio	ons	yes			B\Face_repetition\e		Cancel Ok
	Clusters			1			<pre>B\Face_repetition\e</pre>		
	Status			Pre-clustered			om datasets has cha	an	nged; use STUDY m 🔄 🖻 👢 sub014
	Total size	(Mb))	227.9					⊞
						1			

- 1 we have created a study
- 2 we have generated single trials for this study
 - 3 we made a design = statistical model

4 – now we have to do the stats = estimate the model parameters (b)

 \rightarrow Restrict 'timelim' [-50 650] and use Weighted Least Square

Whilst it's computing, let's recap

- For each subject, at each electrode we have a statistical model: Y = XB + e
- The Weighted Least Square solution is B = pinv(WX)*WY for which the matrix W minimizes the influence of outliers in an otherwise (assumed) normally distributed data-set.
- Outliers are defined in a multivariate space, for instance a trial with a time course 'different' from others
- For each subject we have Yr (data), LIMO (model, weights, other info), Betas (parameters), Yhat (LIMO.design.X*Betas = model), Res (residuals), R2 (model fit), Condition_effect1 (=ANOVA F test across all conditions).



we can mow easily look at each subject

the group level one-sample t-test is also always computed (not always meaningful)

Weighted least squares in LIMO EEG

- Principal Component Projection method:
- PCA
- outlier detection on projected data points (Filzmoser et al., 2008)
- 1 weight per trial



Let's check the weights

• In LIMO Tools, select 'Check weights'



Group level analysis

- Just the same as with ERP but (i) we use betas (ii) LIMO EEG uses robust statistics (essentially 20% trimmed means – except for now the repeated measures ANOVA)
- Call the LIMO GUI, and select random effects
- Need a way to group subject with different channels (no interpolation) – the study created such file for you and it should be loaded by default – if not load:

 \rightarrow limo_gp_level_chanlocs.mat

 Create a folder for the results and do a repeated measure ANOVA selecting the beta files [1 2 3]



Group level analysis

- Call LIMO Results to look at the ANOVA results
- Note the choice between uncorrected and corrected p-values





- If you check the 'course plot' all we have is the contrast used to get the F values
- Best check the 'real' data plotting the betas and ERP → back to Random Effect



Basic stats	- Tests
Central tendency and Cl	One Sample t-test
Plot central tendency and Cl	what dat
Make and plot a difference	type of analysis
Parameter plots	d t-test
- Set Parameters	Regression
Bootstrap 1000 Compute TFCE Analyze ICs	ANOVA/ANCOVA
Working Directory	Load expected chan / neighbours
Help	Quit

Copyright (C) LIMO Team 2015 - GNU GPL

LINEAR MODELING TOOLBOX: Random Effects

limo random effect

Group level analysis

- For Betas select the txt file listing betas of each subject (you can also load one by one if you like clicking)
- For ERP select the txt file listing LIMO files of each subject, use weights
- Using Plot central tendency and CI, you should be able to:
- 1 see the mean beta 1 / 2 / 3 used in the ANOVA
- 2 see the corresponding ERP for Famous faces, scrambled faces and non famous faces

Basic stats	Tests
Central tendency and CI	One Sample t-test
Plot central tendency and Cl	what dat
Make and plot a difference	type of analysis
Parameter plots	dt-test
- Set Parameters	Regression
Bootstrap 1000 Compute TFCE Analyze ICs	ANOVA/ANCOVA
Working Directory	Load expected chan / neighbours
Help	Quit

LINEAR MODELING TOOLBOX: Random Effects

Copyright (C) LIMO Team 2015 - GNU GPL



<u>Our ERP looks a lot like the published results – replicated ! (but not so such betas ?)</u> Bonus: 95% CI are Bayesian = this is the probability of the estimator

Wait a minute ! The model was ...



Y (the data) = FF*Beta1 + SF*Beta2 + NFF*Beta 3 + Beta 4 + e



Let's try faces vs scrambled !

- Per subject compute a contrast famous + non-famous > scrambled
- [1 -2 1] tests [famous 2*scrambled + non-famous] = 0

- At the group level, do a one-sample t-test
- At the group level, you can build an ERP pooling famous + non-famous

LIMO BATCH

faces vs scrambled



Why is LIMO Robust?

- Standard stats are all instantiations of a GLM using an Ordinary Least Square solution → implies looking at the mean
- the breakdown point of an <u>estimator</u> is the proportion of incorrect observations (e.g. arbitrarily large observations) an estimator can handle before giving an incorrect
- For data x1 to xn the mean has a bkdp of 0 because we can make the mean large changing any xi the median has a bkdp of 50%

Why is LIMO Robust?

- Are you sure?
- <u>Micceri (1989). The Unicorn, The Normal Curve, and Other Improbable Creatures.</u> <u>Psych Bul. 105, 156-166</u>
- If the data are Gaussian, the median, the trimmed mean is the same as the mean ! So no reason not to use alternative techniques.
- 1st level, uses weighted least square (weights down bad trials bkdp variable)
- 2nd level involves 20% trimmed mean (weights = 0 for bad subjects): t-tests, 1-way ANOVA, Repeated Measures ANOVA (soon)
- For regressions and N-way ANOVA/ANOVA we use an IRLS (all subjects have weights from 0 to 1 – bkdp variable)

http://en.wikipedia.org/wiki/Robust_statistics

Let's look again at faces vs scrambled



3 x 3 repeated measures ANOVA

- Since we have 3 types of faces and 3 repetition levels we can do a 3 by 3 ANOVA
- Question: How many 1st level regressors?
- Question: Why not modelling interaction at the subject level?

MATLAB R2013a		
HOME PI OTS APPS		CAR A B C C C O Search Documentation
EEGLAB v14.x (dev)	Edit STUDY design pop_studydesign()	
File Edit Tools Plot Study Datasets Help	Select STUDY design	Design Matrix
Study filena Precompute channel measures Study task r Plot channel measures	STUDY.categorical Just me, sorry not in your a	Value
Nb of subjec Precompute component measures Nb of condit PCA clustering (original) Edit/plot clusters	Resave STUDY	<1x3 cell> <1x54 struct> 10 16
Nb of sessic Linear MOdeling EEG Data (LIMO/Components) • Nb of groups 1 per subject Epoch consistency yes	Independent variables New Import Edit Delete Categorical variable: condition - Values (familiar - scrambled - unfamiliar) Sub sub	8 [0.9300,0.9600,1] 010 10
Channels per frame60,65,66,67,68,69,70Channel locationsyesClusters1	Categorical variable: type - Values (13 & 17 & 5 - 14 & 18 & 6 - 15 & 19 & 7)	Add variable Date Modifie Select independent variable
StatusPre-clusteredTotal size (Mb)228	Delete all pre-computed datafiles for this STUDY design Web help Cance	Imt_ume 14/10/2016 2 timedist 14/10/2016 2 trialdist 14/10/2016 2 type 14/10/2016 2 value 14/10/2016 2
<pre>pop_loadset(): loading file E:\My_Documents pop_loadset(): loading file E:\My_Documents pop_loadset(): loading file E:\My_Documents pop_loadset(): loading file E:\My_Documents</pre>	<pre>\MATLAB\Face_repetition\eeg\sub017\ds117_sub017_p05_ \MATLAB\Face_repetition\eeg\sub017\ds117_sub017_p05_ \MATLAB\Face_repetition\eeg\sub017\ds117_sub017_p05_ \MATLAB\Face_repetition\eeg\sub017\ds117_sub017_p05_</pre>	Categorical variable Select variable values
<pre>pop_loadset(): loading file E:\My_Documents pop_loadset(): loading file E:\My_Documents pop_loadset(): loading file E:\My_Documents pop_loadset(): loading file E:\My_Documents</pre>	<pre>\MATLAB\Face_repetition\eeg\sub017\ds117_sub018_p05_ \MATLAB\Face_repetition\eeg\sub018\ds117_sub018_p05_ \MATLAB\Face_repetition\eeg\sub018\ds117_sub018_p05_ \MATLAB\Face_repetition\eeg\sub018\ds117_sub018_p05_</pre>	7 13 & 17 & 5 14 & 18 & 6 15 & 19 & 7 Combine selected values
 ^F ^F ^F Events codes. ^E {'5' '6' '7'}: Familiar faces (1st present) 	entation, 2nd presentation, 3rd presentation delayed)	a Ok

^I { '13' '14' '15' }: NonFamiliar faces (1st presentation, 2nd presentation, 3rd presentation delayed) { '17' '18' '19' }: Scrambled faces (1st presentation, 2nd presentation, 3rd presentation delayed}

3 x 3 repeated measures ANOVA



dataviz @ p = 0.005 uncorrected

Inference TFCE show category effect @ 200ms +, repetition effect @ 600ms +, and interaction at 600ms +



We have a main repetition effect driven by the direct repetition starting circa 380ms but sig only at 550ms

We have a small interaction effect with famous faces showing enhanced activity from circa 550ms

The maths behind the GUI

- If you want to go for coffee now, it's fine
- 4 slides showing how standard and robust are different

One sample t-test



limo_ttest.m

limo_trimci.m

Paired t-test

 $t = \frac{Mean \, (diffeence)}{std \, (difference)/\sqrt{n}}$

limo_ttest.m

$$t = \frac{Difference \ of \ trimmed \ means}{\sqrt{\frac{\left(WinVar1*(n-1)\right) + \left(WinVar2*(n-1)\right) - (2*(n-1)*WinCov)}{(n-2)*n \ trim}}}$$

p = 2 * (1 - tcdf(abs(t), df) with df = ((n - 2) * n trim)-1

limo_yuend_ttest.m

Two-samples t-test



limo_yuen_ttest.m

IRLS

- Limo_wls.m and limo_irls.m (for trials vs across subjects)
- Start by OLS to obtain residuals
- Check outliers in standardized residuals (MAD)
- Compute weights (bisquare function)
- Recompute on weighted data
- Check residuals again until E(e) = 0

→ for eeg, iterate until max(abs(oldRes-newRes)) < (0.0001)

$$Wy = WX \beta + We$$
, $E(e) = 0$, $Cov(e) = \sigma^2 I$