

眼球運動測定による注視時間データと反応時間としての読み時間データの分析

新井 学 (成城大学・東京大学)
m-arai@phiz.c.u-tokyo.ac.jp
Douglas Roland (東京大学)

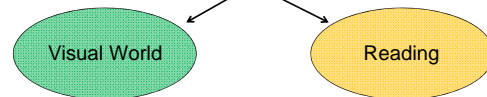
Why is eye-tracking useful for language research?



Eyelink II (SR Research)

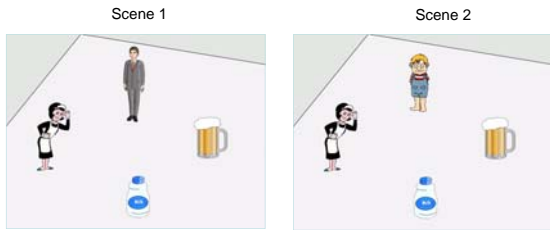
Where readers (listeners) look and how long they look there provides valuable information about the mental processes associated with understanding a given word or set of words (Just & Carpenter, 1980).

Two paradigms



VW experimental design (hypothetical)

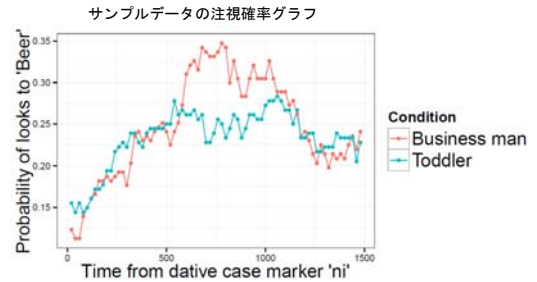
“レストランでウェイトレスが [ビジネスマンに 幼児に] ビールを ミルクを 運んだ”



- 興味対象: 一般世界知識を即座にを使って次に来る項を予想しているか?
- 従属変数: ビール/ミルクのオンセット以前に起きた予測的眼球運動
- 予測変数=1条件2水準 (Scene 1 or Scene 2)

Eye-movement Data

- 眼球運動はジャンプ(Saccade)と停留(Fixation)の連続
- Saccadeは平均20~60 ms: この間に情報はほとんど得られない
- => 主にFixationデータを元に分析が行われる



Time and X-Y coordinates

- オリジナルデータには時間と座標点情報が記録される

Participant	Item	Condition	SegmentStart	SegmentEnd	Recording Time Stamp	GazeEvent Type	FixationPointX	FixationPointY	PupilLeft	PupilRight
3	12	2	288133	293018	288307	Saccade			3.66	3.54
3	12	2	288133	293018	288308	Saccade			3.66	3.55
3	12	2	288133	293018	288309	Fixation	467	845	3.67	3.55
3	12	2	288133	293018	288310	Fixation	467	845	3.67	3.55
3	12	2	288133	293018	288311	Fixation	467	845	3.67	3.53
3	12	2	288133	293018	288312	Fixation	467	845	3.67	3.53
3	12	2	288133	293018	288313	Fixation	467	845	3.66	3.54
3	12	2	288133	293018	288314	Fixation	467	845	3.66	3.53
3	12	2	288133	293018	288315	Fixation	467	845	3.65	3.54
3	12	2	288133	293018	288316	Fixation	467	845	3.65	3.53
3	12	2	288133	293018	288317	Fixation	467	845	3.65	3.53
3	12	2	288133	293018	288318	Fixation	467	845	3.65	3.53
3	12	2	288133	293018	288319	Fixation	467	845	3.65	3.54
3	12	2	288133	293018	288320	Fixation	467	845	3.65	3.52
3	12	2	288133	293018	288321	Fixation	467	845	3.63	3.53
3	12	2	288133	293018	288322	Fixation	467	845	3.64	3.54
3	12	2	288133	293018	288323	Fixation	467	845	3.64	3.52
3	12	2	288133	293018	288324	Fixation	467	845	3.64	3.52
3	12	2	288133	293018	288325	Fixation	467	845	3.63	3.54
3	12	2	288133	293018	288326	Fixation	467	845	3.64	3.54
3	12	2	288133	293018	288327	Fixation	467	845	3.62	3.53
3	12	2	288133	293018	288328	Fixation	467	845	3.62	3.54
3	12	2	288133	293018	288329	Fixation	467	845	3.62	3.52
3	12	2	288133	293018	288330	Fixation	467	845	3.62	3.52

Mapping and Converting

- 座標軸情報を視覚刺激上の各エリアにマッピングする



- データは各エリアに対して0 (On)か1 (Off)の二項変数になる

Participant	Item	Condition	SegmentStart	SegmentEnd	Recording Time Stamp	GazeEvent Type	FixationPointX	FixationPointY	ACH	ACR
3	12	2	288133	293018	288347	Saccade			0	0
3	12	2	288133	293018	288350	Saccade			0	0
3	12	2	288133	293018	288354	Fixation	467	845	1	0
3	12	2	288133	293018	288357	Fixation	467	845	1	0
3	12	2	288133	293018	288360	Fixation	467	845	1	0
3	12	2	288133	293018	288364	Fixation	467	845	1	0
3	12	2	288133	293018	288367	Fixation	467	845	1	0
3	12	2	288133	293018	288370	Fixation	467	845	1	0
3	12	2	288133	293018	288374	Fixation	467	845	1	0
3	12	2	288133	293018	288377	Fixation	467	845	1	0
3	12	2	288133	293018	288379	Fixation	467	845	1	0
3	12	2	288133	293018	288380	Fixation	467	845	1	0
3	12	2	288133	293018	288383	Fixation	467	845	1	0
3	12	2	288133	293018	288386	Fixation	467	845	1	0
3	12	2	288133	293018	288389	Fixation	467	845	1	0
3	12	2	288133	293018	288394	Fixation	467	845	1	0
3	12	2	288133	293018	288397	Fixation	467	845	1	0

Choosing time window

- 試行全体、もしくはある一点のサンプリングデータの眼球運動パターンを予測するのは不可能
- => 興味ある特定の時間幅を選んで「特定の対象に対してどれだけの注視量があったか」を見る



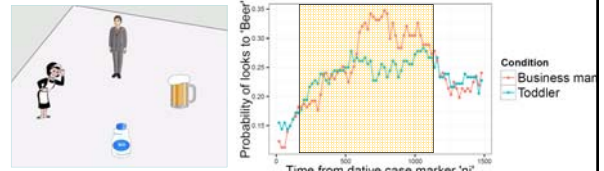
- 各言語刺激内の語彙オンセットを元に時間枠を決める
- e.g., 「に」のオンセット+200 msから「ビール/ミルク」のオンセットまで



NB: この場合時間枠の長さは各項目によって異なる

Choosing time window

- 例えば 1200 ms の時間幅中で600個のデータポイントがあった時に、300個ビールを見て、100個ミルクを見て、100個ウェイトレスを見て、100個どこも見ていなかった(欠損)とする。



=> この時間枠内で50%ビールを注視していたことになる

Not probability but logit

- 二項分布の平均確率は線形モデルには不適切 (e.g., Jaeger, 2008)
- => ロジット(logit)に線型モデルを仮定する

$$\ln \frac{p}{1-p} = \beta_0 + \beta_1 x$$

- 便宜上分母分子に0.5を足したempirical logitがよく用いられる

$$\eta' = \ln \frac{(y+0.5)}{(n-y+0.5)}$$

e.g., $\ln \frac{300+0.5}{500-300+0.5}$ NB: 欠損値を含めない

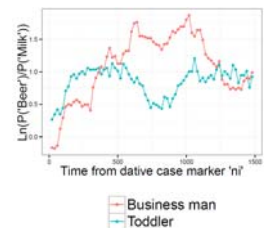
=> 「注視全体からしてどのくらいの比率でビールを見ていたか」

Log-ratio

- 時として興味対象は二つの対象物への注視のみ（「ビール」 vs. 「ミルク」）
- 二つの対象物間の注視の比率 (ratio) を対数変換する
- n をビールとミルクへの注視量の総和と考えることと同じ

$$\eta' = \ln \frac{(y+0.5)}{(n-y+0.5)}$$

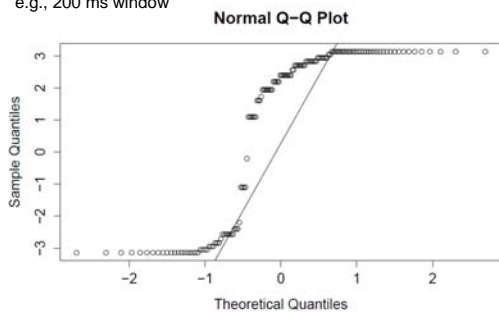
e.g., $\ln \frac{300+0.5}{100+0.5}$



Window size issue

- 時間枠が小さすぎるとデータはほぼ0か1に近くなる (また欠損のみのinvalid dataが増える)

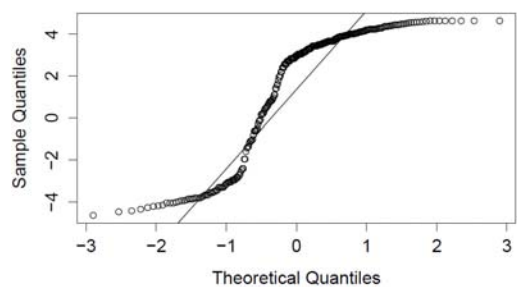
e.g., 200 ms window



Window size issue

1000 ms window

Normal Q-Q Plot

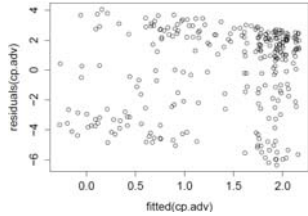


➢ たゞ時間枠が大きいても正規分布に当てはまらないケースがほとんど

Residual Analysis

- しかし一般化線形混合モデルにおいてはモデルの残差が正規分布に従っていれば従属変数自体が正規分布に従う必要はない
- => 残差の分布をチェックする

Residuals vs. fits plot



Alternative 1 – Modelling changes over time

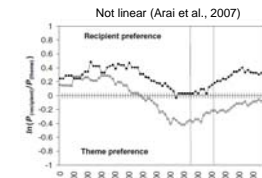
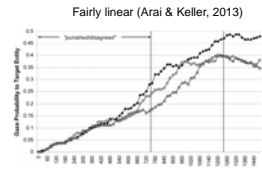
- 連続する複数の時間枠 (50 msとか 100 ms) を決め、その順番を変数として説明変数 (Time) に加える

=> 時間軸に沿った注視量の変化を見ることができる。



- 各時間枠は短くなる
- 時系変化は線形でない場合が多い

=> Quadratic term (Time²)をモデルに加える
=> モデルが複雑になる



Alternative 2 – Permutation Analysis

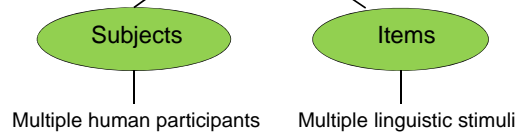
- 小さな連続する各時間枠に対してtest statisticを計算し、有意となった連続するbinをclusterとしてまとめる。そのClusterに対し説明変数のラベルをランダムにscrambleしtest statisticを計算することを繰り返し (permutation)、実際にデータから得られたtest statisticsがpermutationから得られたtest distributionに比べて有意に小さいかチェックする、というノンパラメトリック分析法。

- ◆ メリット

- 実験者が恣意的に時間枠を選ばない
- Multiple comparisonの問題を回避
- 隣接する時間枠の依存関係の問題はclusterを作ることで回避

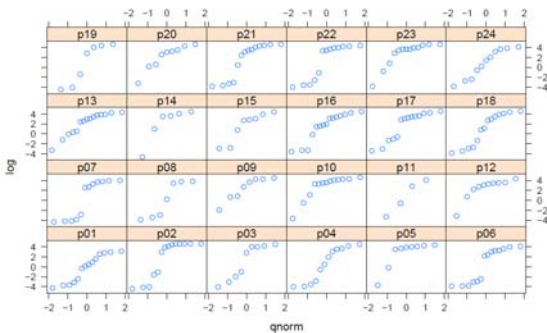
Sources of variance:

Random effects



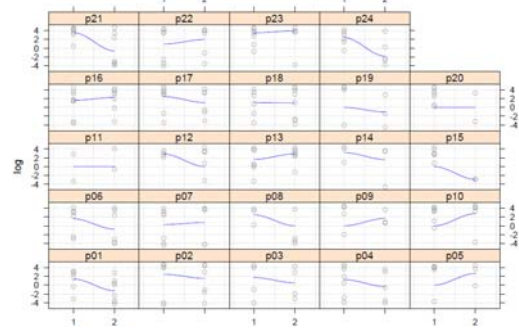
1. By-subject random intercept

- 被験者毎に分布が異なる



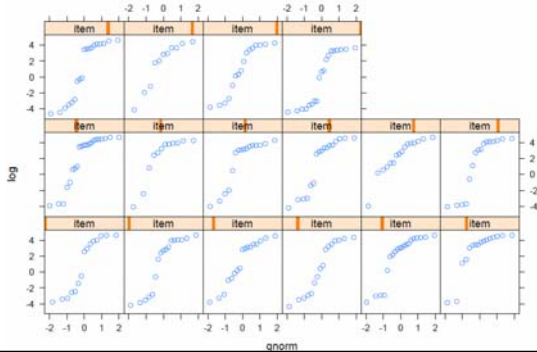
2. By-subject random slope

- 被験者毎に予測変数の影響度が異なる



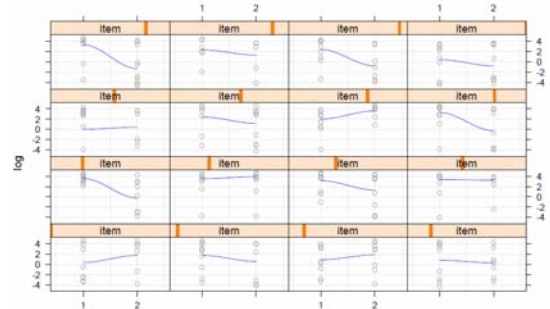
3. By-item random intercept

- アイテム毎に分布が異なる



4. By-item random slope

- アイテム毎に予測変数の影響度が異なる



=> Language is not a fixed effect (Clark, 1973).

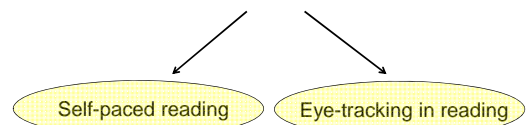
Advantages of Linear Mixed-Effects Models for psycholinguistic experiments

- 分散分析を代表とする今までの統計手法では被験者、又はアイテム毎の平均値を求める (Averaging) 必要があった (F1 & F2分析)
 - 線形混合モデルではそのAveragingを必要とせず、各トライアルのデータをそのまま従属変数として使える
- => 各実験の流れの中で説明変数の効果がどのように起こったか (または変化したか) 知ることができる
- => 隣接するトライアルが説明変数の効果にどう影響を与えたかなど、今まで無視されていたが関係があったかも知れない共変数 (covariate) の影響を調査できる



(Barr, 2008)

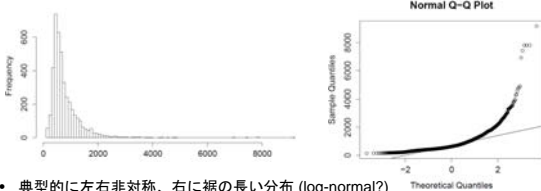
Reading time data



文節に区切られた刺激に対する反応時間 自然な読みの中での注視時間
(後で文節に区切って分析する)

Self-paced reading data

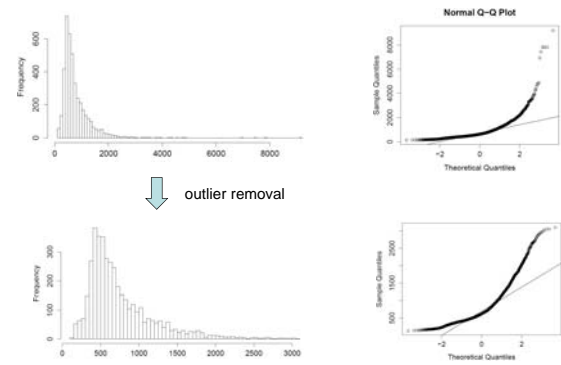
- あるリージョン (語や句) に対する反応時間を読み時間の指標とする



- 典型的に左右非対称、右に裾の長い分布 (log-normal?)
- 自己ペース読み実験の読み時間データは一般的に
 - (1) エラーによるOutlierを取り除く Data trimming
 - (2) 極端な値の影響を弱める為SD値を基準にBoundary Value Replacementが行われる
- * 正規分布を仮定せずモデルの残差を使う方法もある (Baayen et al., 2008)
- (フィルターを含めた) データ全体を使って、被験者間のベースパフォーマンスの差異、リージョンの影響、文字数の影響などを先に取り除いてその残差読み時間を従属変数とする手法も使われる (cf. Ferreira & Clifton, 1986 ; Jaeger lab blog)

Self-paced reading data

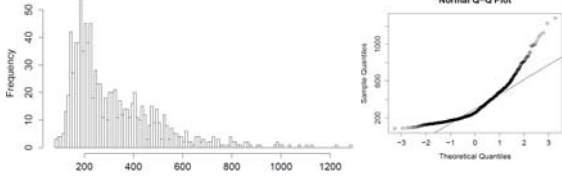
- Outlier removalはデータの分布に大きく影響する



Eye-tracking reading data

- あるリージョン（語や句）におけるFixation（の長さ）を指標にする

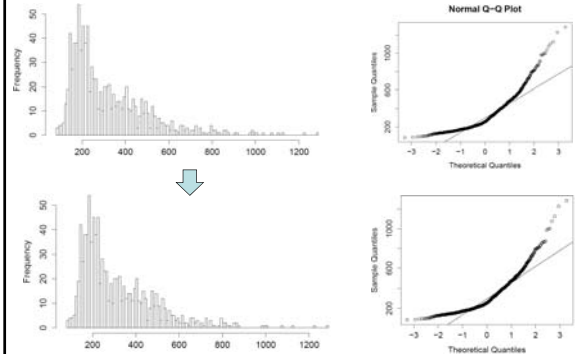
あるリージョンのFirst pass times



- 典型的に左右非対称、右に裾の長い分布
- 眼球読みデータにおいては一般的にData trimmingや残差読み時間の計算は行われない（しかしやってはいけない理由はない）
- 一般的にSD値を元にBoundary Value Replacementのみが行われる（又はモデルの残差を使う）

Eye-tracking reading data

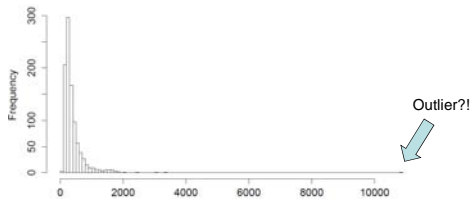
- Boundary Value Replacementは分布にほぼ影響しない（3SD以上を除いた例）



No outlier in eye-tracking reading data?

- 眼球読み実験においてはエラーによるoutlierは存在しないか？

ある実験の一つのリージョンのRegression path times

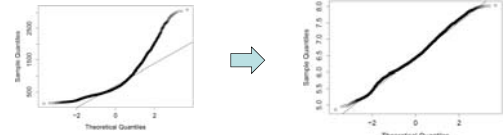


- ヒューマンエラーはどんな実験であれ起こりうる
- 眼球読み実験においては特に読み返しを含む指標に注意

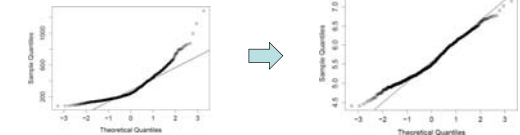
Should we log-transform data?

- VW eye-tracking data => 元が二項変数 => Yes
- Reading data => 元が連続変数だが正規分布に従わないことが多い => Possibly yes but be careful

> SPR times



> Eye-tracking reading times



(See Frank, Monsalve, Thompson, & Vigliocco, 2013)

Should we perform model selection?

選択肢① Maximal LMEMsを常に使う

=> Barr (2013), 'Random effects structure for confirmatory hypothesis testing: Keep it maximal'

=> 実際のデータ・デザインにそぐわないケースがある

選択肢② Maximal LMEMsから始め最適なRandom effect structureを持つ最良なモデルを選択する

=> Principle of parsimony (Occam's razor)

=> Bates et al. (submitted), 'Parsimonious mixed models'

"Overparameterization can lead to uninterpretable models"

サンプルアウトプット

```
Linear mixed model fit by REML
Formula: log ~ DativeN + (1 + DativeN | subj)
          + (1 + DativeN | item)
Data: cp_ord.cast
AIC: BIC loglik deviance REMLdev
1369 1401 -675.5 1350 1351
Random effects:
Groups Name Variance Std.Dev. Corr
subj (Intercept) 0.117606 0.34294
      DativeN 0.493045 0.70217 (1.000)
item (Intercept) 0.043113 0.20764
      DativeN 0.860472 0.92762 0.768
Residual 8.495560 2.91471
Number of obs: 269, groups: subj, 24; item, 16

Fixed effects:
Estimate Std. Error t value
(Intercept) 1.3924 0.1999 6.964
DativeN -1.1342 0.4533 -2.502

Correlation of Fixed Effects:
(Intr)
DativeN 0.231
```

Summary & Conclusion

- ✓ 眼球運動測定実験で得られるデータはパラダイムによって性質が異なり、予測変数の影響は適切な従属変数なしには計れない
- ✓ ランダム効果には被験者に起因するものと、項目に起因するもの、予測変数と独立しているもの、関係するものがある
 - 通常被験者によるランダム効果は統制不可能で大きい
- ✓ 混合線形モデルは複数のランダム効果をaveragingなしでモデルに同時に含めることができる
- ✓ 個々のトライアルのデータを扱うことで、予測変数の影響の時間変化や、様々な共変数の影響を調べられる
- ✓ しかしデータの分布、Outlierの存在、Boundary valueの影響には注意
- ✓ 元々二項変数のVisual World眼球データではlogitを計算して混合モデルが適応される（一般化線形混合モデル）
- ✓ 元々連続変数の読み時間データは対数変換をすると通常正規分布に近似するが、常に変換すべきか否かはまだ不明
- ✓ Overparameterizationの問題を避ける為モデル選択は(おそらく)するべき