Statistics of FMRI

Gang Chen

Scientific and Statistical Computing Core
NIMH/NIH

August 15, 2017

Overview of FMRI Statistics

❑ What is statistics? What is statistical inference?

❑ Designed experiment (vs observational study)

  ✧ How to efficiently squeeze the signal of interest into the data

❑ Typical statistical models involved in FMRI

  ✧ Regression: individual level

  ✧ General linear model (GLM): group level

    ▪ Student's t-tests: one-, two-sample, or paired

    ▪ Multiple regression

    ▪ AN(C)OVA

  ❑ Linear mixed-effects (LME) model

❑ Miscellaneous

What is statistics?

- In God we trust. All others must bring **data**!
  - Related to *state* and *status*: science of state
  - German *statistik* by Gottfried Achenwall (1749)
- Statistics is about dealing with variation and noise
  - Variation: the world is not uniform
  - Noise is annoying, but contains a lot of useful information
  - Art of noise handling
- Science of data analysis based on variation
  - Experimental design and data collection
  - Model building, comparison, selection and analysis
  - Inference and interpretation
  - Presentation and prediction

Is our brain statistically wired?

- ❑ Example 1
  - ➢ Should we buy lottery tickets and insurances?
- ❑ Example 2
  - ➢ Average deaths due to snakebites in the USA and Canada per year? *15*
  - ➢ Average deaths in air transport per year in the USA? *200*
  - ➢ Average number of people killed by cars annually in the USA? *40,000*
- ❑ Example 3
  - ➢ Suppose I flip a coin 5 times, and get all heads. What is the chance of getting a head in the next toss?
- ❑ Example 4
  - ➢ HIV prevalence = 0.1%, false + of HIV test = 5%, power of HIV test ~ 100%. P(HIV+ | test+) = ?

$$P(HIV+|test+) = \frac{P(test+,HIV+)}{P(test+)} = \frac{P(test+|HIV+)P(HIV+)}{P(test+|HIV+)P(HIV+) + P(test+|HIV-)P(HIV-)} = \frac{1.0 \times 10^{-3}}{1.0 \times 10^{-3} + 0.05 \times (1 - 10^{-3})} \approx 0.02$$

Statistics in daily life

- Census in Bible
- Stereotype: economical thinking
- Miss a step when walking downstairs
- Lottery and gambling
  - Dream of a huge gain out of an unlikely occurrence by paying > expected value
- Insurance
  - Risk management: hedge the risk of contingency losses
  - Opposite of lottery: Avoid a huge incidental loss by paying > expected cost
  - Airbags, ABS, alarm systems, security monitoring system…
- Surveys, clinical trials, …

Experiment Designs

- FMRI experiment types
  - Task-based
    - Block design
    - Event-related
    - Mixed type
  - Resting state

  - Naturalistic FMRI
    - Movie watching, music/speech listening

Experiment Design

❑ **Principle:** *Jam all possible information into the design before data acquisition*

❑ Factors for an efficient design at the individual level

- ✧ Number of time points (TRs)
  - ▪ Important for individual analysis, but may affect group level implicitly
  - ▪ Power: proportional to $\sqrt{DF}$
  - ▪ Limited by subject's tolerance in scanner: 30-90 min per session
- ✧ TR length: mostly 2 sec
  - ▪ Shorter TR yields more time points (and potentially more power), but
  - ▪ Power improvement limited by weaker MR signal (SNR)
  - ▪ Shorter TR → shorter ISI → higher event freq → higher correlation → less power
  - ▪ Usually limited by hardware considerations

Experiment Design: individual level

❑ Factors for an efficient design at the individual level

- ✧ Complexity of the experiment: number of conditions/tasks (regressors)
  - ▪ Limited by scanning time and confounded by low frequencies
- ✧ Sample size: number of trials per condition/task
  - ▪ The more the better, but no magic number: block or event-related?
- ✧ Event arrangement
  - ▪ How to design? How to define the **best** design?
  - ▪ Efficiency: achieve highest statistical power within fixed scanning time
  - ▪ Programs for randomizing trials

Experiment Design: individual level

❑ Factors for an efficient design at the individual level

✧ HDR modeling

▪ Fixed-Shape Method (FSM): assuming an empirical response curve; economical but high risk of inaccuracy

▪ Adjusted-Shape Method (ASM): more wiggle room for shape variability; focus on the major shape

▪ Estimated-Shape Method (ESM): capable of achieve accurate shape characterization and power; challenging on group analysis
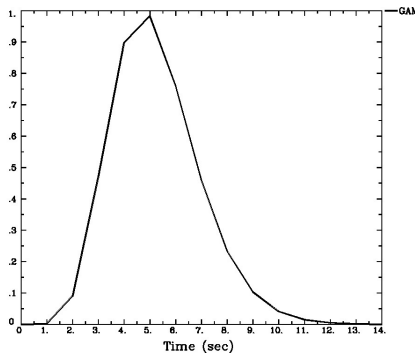
$$h(t) = t^{8.6} \exp(-t/0.547)$$

Experiment Design: group level

- ❑ Factors for an efficient design at the group level
    - ✦ Number of subjects (*n*)
        - ▪ Important for group analysis: inter-subject vs. intra-subject variation
        - ▪ Power (success to detect signal if present) roughly proportional to $\sqrt{n}$
        - ▪ Design type: block vs. event-related
        - ▪ Recommended: 20+
    - ✦ Design of the study
        - ▪ Complexity: factors, levels, covariate, contrasts of interest, …
        - ▪ With multiple groups, counterbalance, if possible, potential confounding effects such as age, gender, IQ, education background, socioeconomic status, etc.
- ❑ At the end of the day
    - ✦ Group level matters the most!

Modeling

- Data visualization: **trust your eyes more than software**
    - Any excessive head motion? Spikes? Abnormalities?
- Modeling building/selection
    - Individual level: time series regression
        - Regressors of interest
        - Regressors of no interest
        - Noise
    - Group level
        - t-tests, Regression, ANOVA, ANCOVA, GLM, LME, …
        - Nonparametric methods
            - Ranking methods
            - Permutations
            - Bootstrapping

Overview: Individual subject analysis

❑ Basics of linear regression model

❑ FMRI experiment types

  ✧ Block design; Event related experiment; Mixed

❑ FMRI data decomposition: three components

  ✧ Effects of no interest: baseline, slow drift, others; Effects of interest; *Noise*

  ✧ Effects of interest - understanding BOLD vs. stimulus: IRF

❑ 3 modeling strategies

  ✧ Model-based: presumed or fixed-shape IRF (FSM)

  ✧ Data-driven: no assumption about IRF shape (ESM)

  ✧ Intermediate: one major IRF plus shape adjustment (ASM)

Basics of Regression

❑ Statistical modeling (information extraction)
  ✧ Two goals
    ▪ Prediction: machine learning – support vector machine (SVM)
    ▪ Inferences: activation detection

❑ Regression: relationship between a response/outcome (dependent) variable and one or more explanatory (independent) variables (regressors)
  ✧ Simple regression - fit data with a straight line: Sir Francis Galton's original meaning - regression to mean
    ▪ When 2 variables are not perfectly correlated, regression to mean exists
    ▪ Psychology (Daniel Kahneman): Rewards for good performance vs. punishment of mistakes (correlation vs. causation)
    ▪ Lost in most cases including FMRI
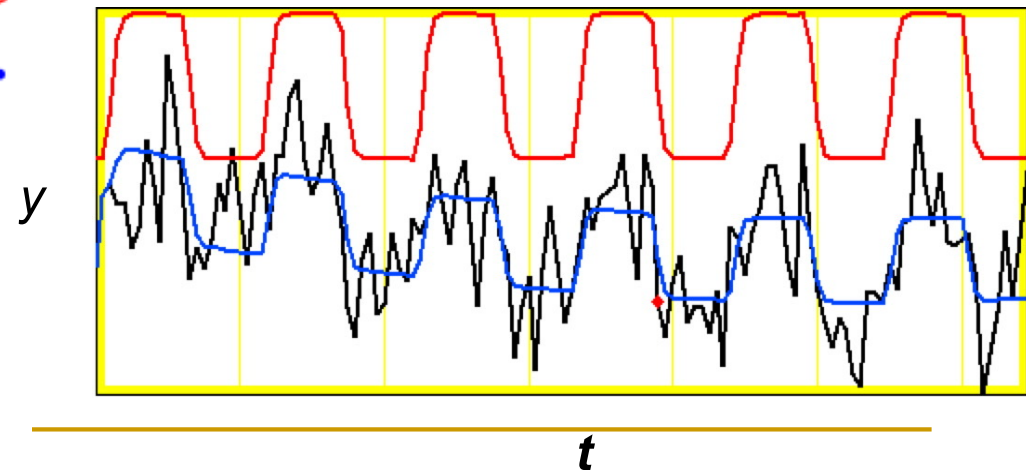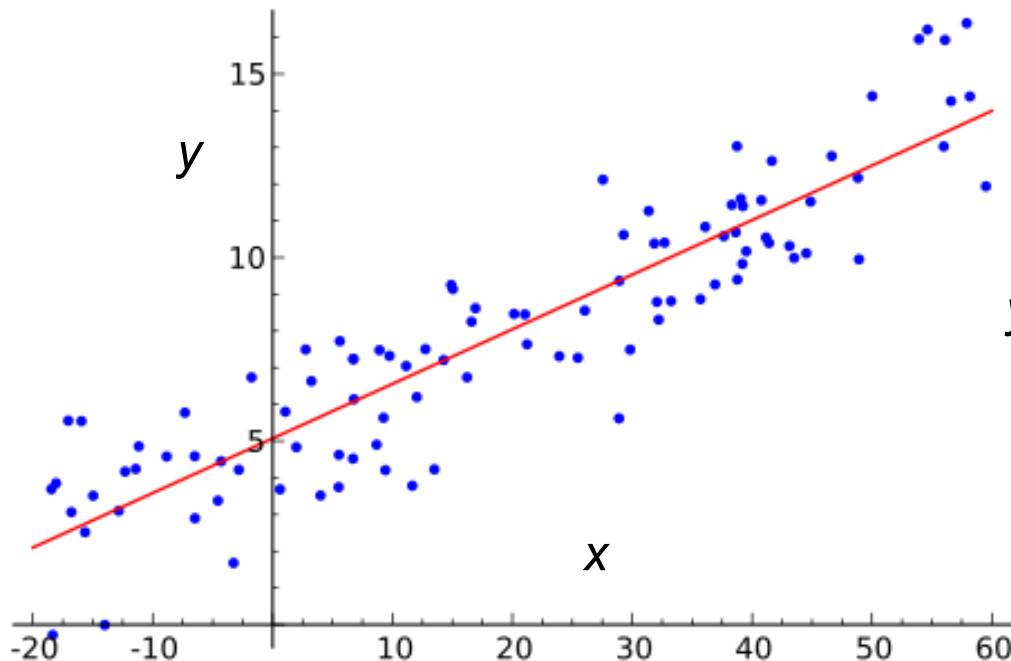 ✧ Some statisticians call it (general) linear model

Basics of Regression

❑ Mathematical formulation

✧ $y_i = \alpha + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} + \varepsilon_i$, i: time index

✧ $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, $\mathbf{X} = [1, x_1, x_2, \ldots, x_k]$

✧ Assumptions

- linearity
- white noise (independence) and Gaussianity $\boldsymbol{\varepsilon} \sim N(0, \sigma^2\mathbf{I})$

Basics of Regression

- Solution for regression $y = X\beta + \varepsilon$
  - ✧ Project data **y** onto the space of explanatory variables (**X**)
  - ✧ OLS $\hat{\beta} = (X^T X)^{-1} X^T y$
- Interpreting β values: slope, marginal effect, or effect estimate associated with a regressor (explanatory variable)
- Various statistical tests
  - ✧ t-test for each β $(H_0: \beta_{sad} = 0)$
  - ✧ t-test for linear combination of β values - general linear test (GLT), *e.g.*, $H_0: \beta_{hap} - \beta_{sad} = 0$, or $H_0: 0.5^*(\beta_{hap} + \beta_{sad}) - \beta_{neu} = 0$
  - ✧ F-test for composite null hypothesis, *e.g.*, $H_0: \beta_{hap} = \beta_{sad} = \beta_{neu}$ or $H_0: \beta_{hap} = \beta_{sad} = \beta_{neu} = 0$
  - ✧ Omnibus or overall F-test for the **whole** model, *e.g*, H_0: all $\beta's = 0$, or for a **partial** model H_0: all $\beta's$ of interest $= 0$

Regression with FMRI

❏ Time series regression: data y is time series

  ✧ Regressors: idealized responses or basis functions

  ✧ Special handling: noise not white $\varepsilon \sim N(0, \sigma^2 \Sigma)$, but with temporal or serial correlation

    ▪ Banded variance-covariance matrix Σ

  ✧ AKA general linear model (GLM) in other FMRI packages

    ▪ General vs. generalized

❏ Same model for all voxels in the brain

  ✧ Simultaneously solve the models: voxel-wise analysis, massively univariate method

  ✧ $y = X\beta + \varepsilon$: same design matrix X across the brain

FMRI Data

❑ Data partition: **Data = Signal + Noise**

    ✦ <u>**Data**</u> = acquisition from scanner (voxel-wise time series)

    ✦ <u>**Signal**</u> = BOLD response to stimulus; effects of interest + no interest

        ■ We don't fully know the real signal!

        ■ Look for idealized components, or search for signal via repeated trials

        ■ Of interest: effect estimate (response amplitude) for each condition

        ■ Of no interest: baseline, slow drift, head motion effects, …

    ✦ <u>**Noise**</u> = components in data that interfere with signal

        ■ Practically the part we have don't know and/or we don't care about, or the part we can't explain in the model

        ■ Will have to make some assumptions about its distribution

❑ **Data = baseline + slow drift + other effects of no interest + response$_1$ + … + response$_k$ + noise**

    ✦ How to handle the effects of no interest?

    ✦ How to construct the regressors of interest (responses)?

    ✦ Assumptions about the noise?

Effects of interest

❑ data = effects of no interest+response$_1$+…+response$_k$ + noise

❑ 3 components: baseline + drift + other effects of no interest

  ✧ Baseline (and drift) can be modeled as an additive effect (AFNI), or an effect of interest (*cf.* SPM and FSL)

  ✧ Drift: psychological/physiological effect + thermal fluctuation; modeled with polynomials

  ✧ $y_i = \alpha_0 + \alpha_1\,t_i + \alpha_1\,t_i^2 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} + \ldots + \varepsilon_i$

  ✧ $y = \mathbf{X}\boldsymbol{\beta} + \varepsilon,\ \mathbf{X} = [1,\ t,\ t^2,\ x_1,\ x_2,\ \ldots,\ x_k,\ \ldots]$

  ✧ Other effects of no interest: head motion effects, censored time points, …

Constructing Regressors: FSM

❑ Assuming a <u>fixed-shape</u> $h(t)$ for HDR to an instantaneous stimulus: impulse response function (IRF)

   ✧ GAM(p,q): $h(t) = [t/(p*q)]^p * exp(p-t/q)$

      ▪ DA variation: SPM (undershoot) - canonical

      ▪ Default IRF: $h(t) = t^{8.6} \exp(-t/0.547)$ [MS Cohen, 1997]

   ✧ Build HDF based on presumed IRF through convolution

      ▪ Roll IRF $h(t)$ with stimulus timing $S(t)$: $x(t) = h(t) \otimes S(t)$
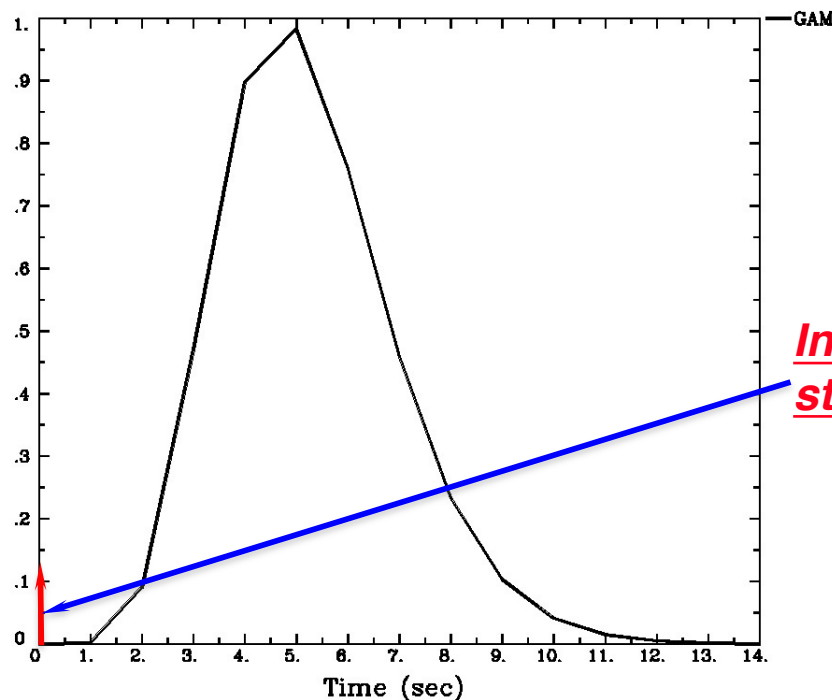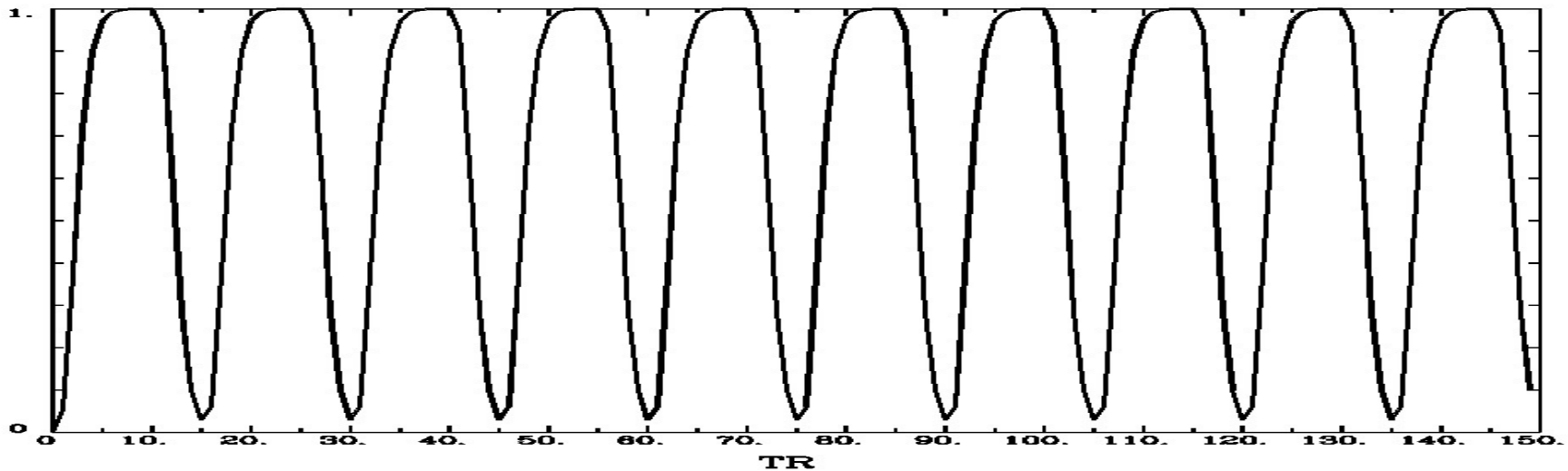

Instant stimulus
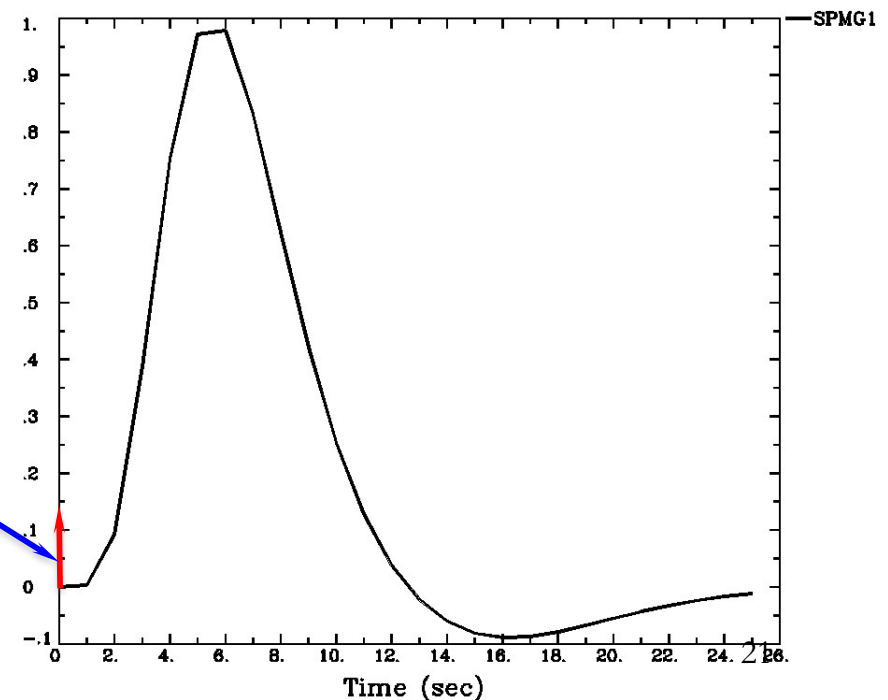
FSM for Block Design

❑ Presuming a <u>fixed shape</u> IRF *h(t)* for each instantaneous stimulus – 3 parameters: onset, duration, peak

  ✦ For each block, *h(t)* is convolved with stimulus timing and duration (*d*) to get idealized response (temporal pattern) as an explanatory variable (regressor): BLOCK(*d,p*)

  ▪ Equivalent to convolving a series of consecutive events
  ▪ Linearity assumed within each block: plateau-like response
  ▪ *p*: scale HDF to 1 for easy interpretation of β

Block: 10 s on and 10 s off; TR=2 s; 150 time points

FSM for Event-Related Design

❑ <u>Fixed shape</u> IRF $h(t)$ for an instantaneous stimulus – 2 parameters: timing and peak

✦ For multiple trials of a condition/task, $h(t)$ is convolved with stimulus timing to get idealized response (temporal pattern) as an explanatory variable (regressor): GAM(p,q) or BLOCK(0)
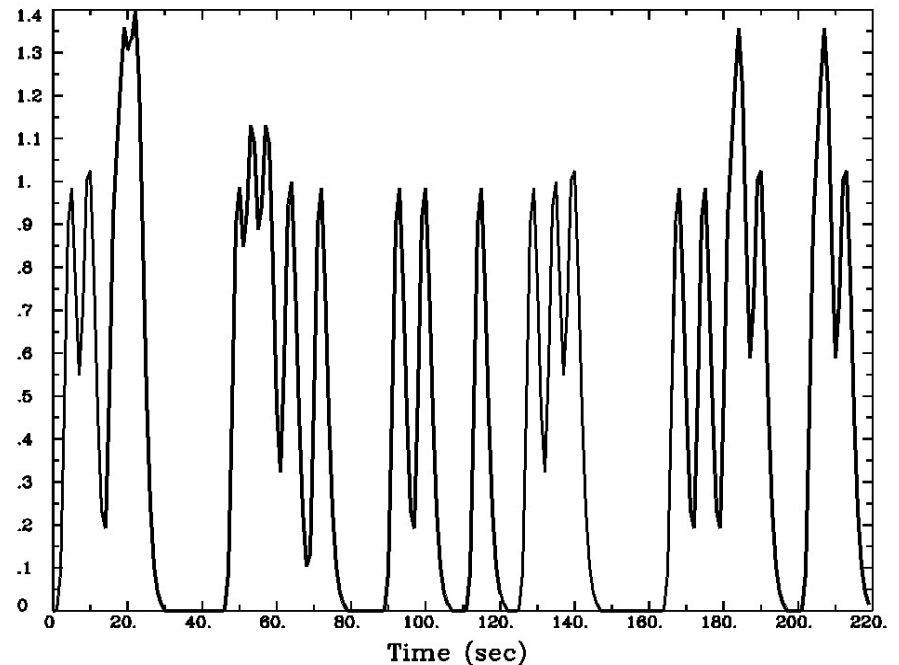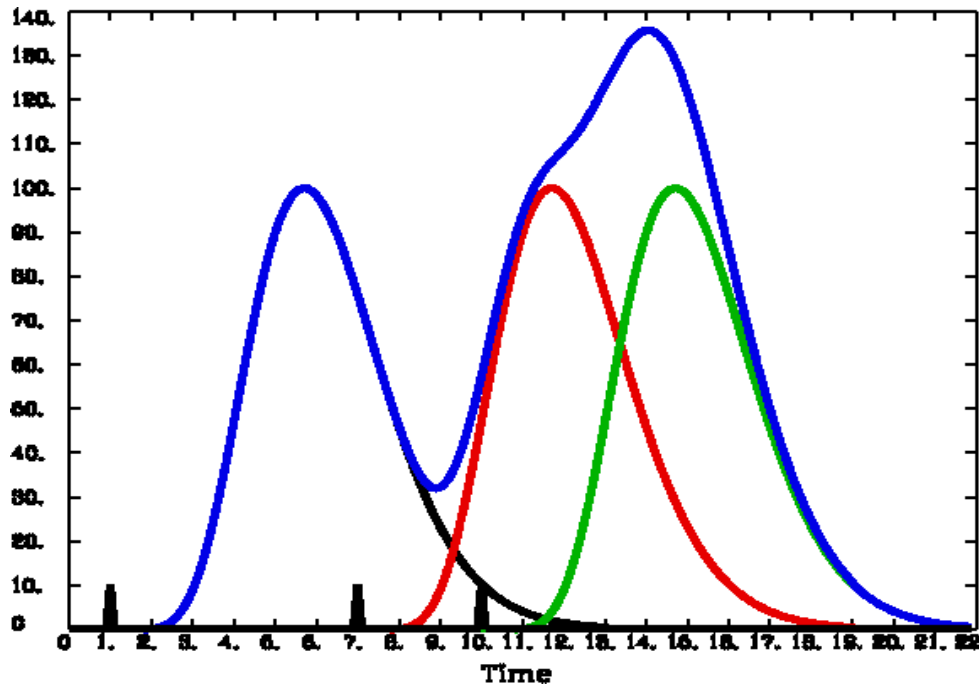
Assessing FSM

❑ Used 99% of time, but not necessarily the optimal

　✦ Assumes brain responds with **same shape** across 4 levels: subjects, activated regions, stimulus conditions/tasks, trials

　　▪ Difference in magnitude β and its significance

　　▪ Strong assumption about 4 levels regarding shape

　✦ Easy to handle: one value per effect

　✦ Works relatively well

　　▪ Block design: shape usually not important due to accumulating effects (modeled via convolution) of consecutive events

　　　▪ Really plateau? Same magnitude across blocks?

　　▪ Event-related experiment

　　　▪ Linearity when two responses overlap? Same effect across events?

❑ Not desirable if you

　✦ care about subtle shape difference across subjects, across regions, across conditions, and across trials

　✦ improve modeling

ESM: No Constraint on IRF Shape

❑ Yardstick (or TENT) perspective
  ✧ Set multiple yardsticks (or tents) at various equally-spaced locations to cover the potential BOLD response period
    ▪ Each yardstick or TENT is a basis function
  ✧ BOLD response measured by yardstick heights at all locations
    ▪ Condition effect is reflected by as many as number of yardsticks
❑ Yardsticks (percent signal change sticks): TENT functions
  ✧ Also known as 'piecewise linear splines'

yardstick with unit height
@ location 3•TR

$$T(x) = \begin{cases} 1 - |x| & \text{for } -1 < x < 1 \\ 0 & \text{for } |x| > 1 \end{cases}$$

$$T\left(\frac{t - 3 \cdot TR}{2 \cdot TR} \right)$$

h

Cubic splines
are also available

time

t=0   t=TR   t=2•TR t=3•TR t=4•TR t=5•TR

25
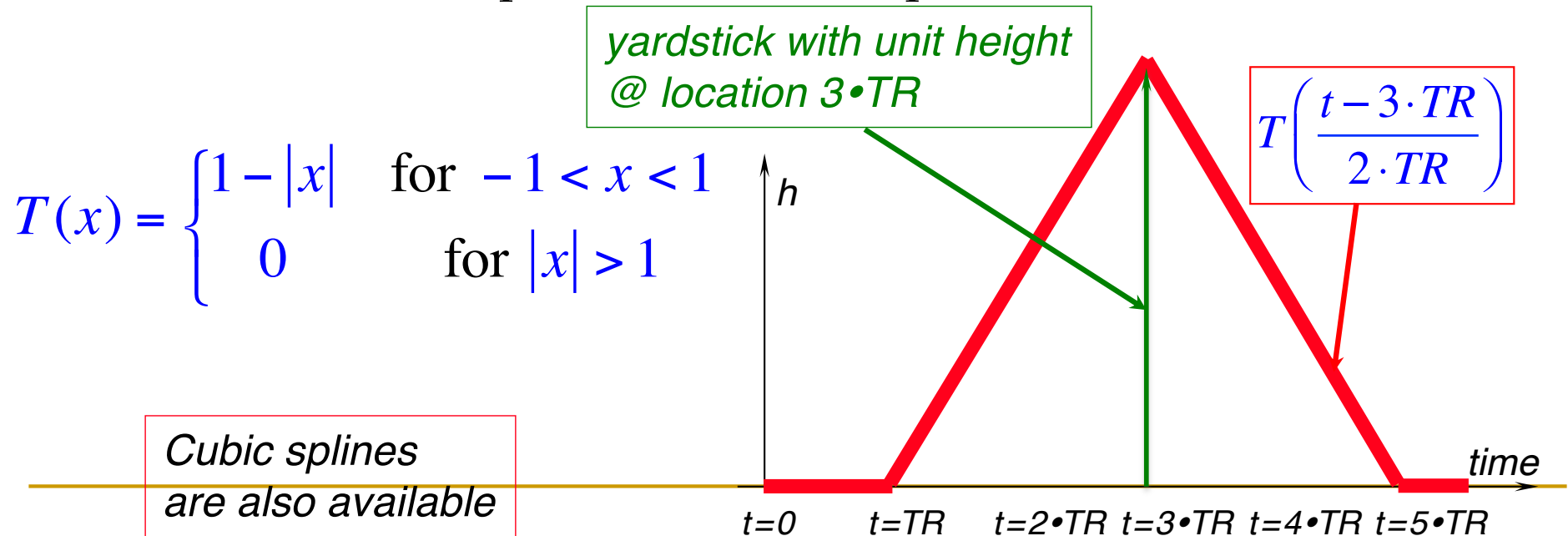
Tent Functions = Linear Interpolation

- 5 equally-spaced tent functions (yardsticks): linear interpolation between "knots" with TENTzero(b,c,n) = TENTzero(0,12,7)

$$h(t) = \beta_1 \cdot T\left(\frac{t-L}{L}\right) + \beta_2 \cdot T\left(\frac{t-2 \cdot L}{L}\right) + \cdots + \beta_5 \cdot T\left(\frac{t-5 \cdot L}{L}\right)$$

CSPLIN(b,c,n): smoother

6 intervals = 5 β weights

stimulus onset

"knot" times
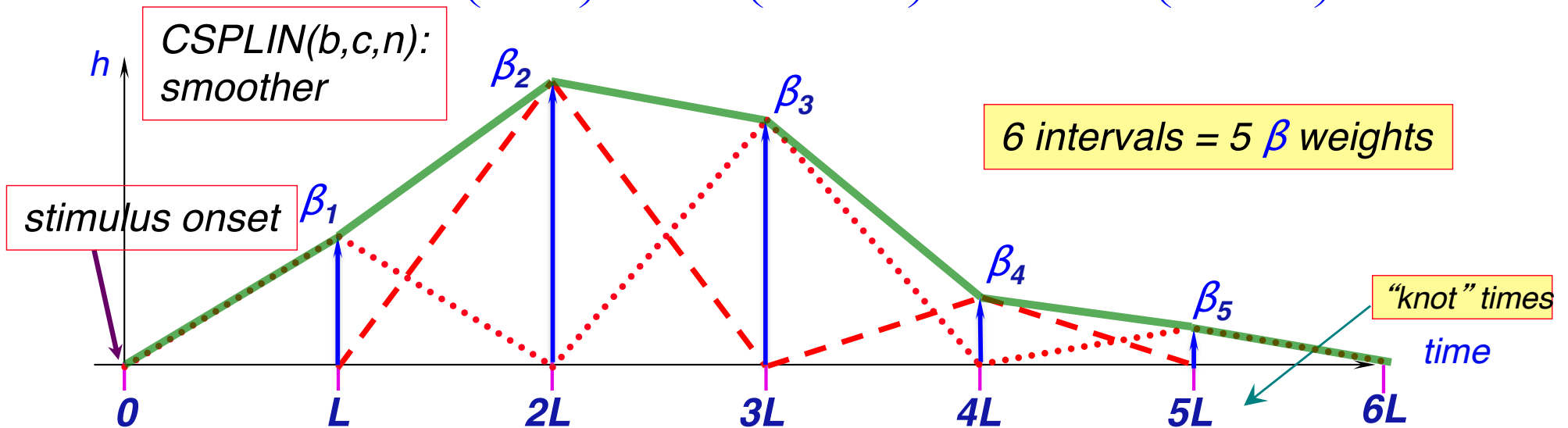
- Tent parameters are easily interpreted as function values (e.g., **L**: tent radius; β_2 = response (tent height) at time $t = 2L$ after stimulus onset)
- Relationship of tent spacing **L** and TR (**L** $\geq$ TR), *e.g.*, with TR=2s, **L**=2, 4s

Assessing ESM

❑ Pros

  ✧ Usually for event-related experiments, but can be used for BLOCK
    ▪ Multiple basis functions for blocks: within-block attenuation
    ▪ Cross-block attenuation?
  ✧ Likely to have more accurate estimate on HDR shape across subjects, conditions/tasks, brain regions
  ✧ Likely to have better model fit
  ✧ Likely to be statistically more powerful

❑ Cons

  ✧ Difficult to summarize at group level: group analysis capability
  ✧ A few times more regressors than alternatives: DF's
  ✧ Risk of highly correlated regressors: multicollinearity
    ▪ Try changing the number of basis functions
  ✧ Over-fitting: picking up something (head motion) unrelated to HDR

Why Group Analysis?

- ❑ Reproducibility and generalization
  - ✧ Science strives for generality: summarizing subject results
  - ✧ Typically 10 or more subjects per group
  - ✧ Exceptions: pre-surgical planning, lie detection, …
- ❑ Why not one analysis with a mega model for all subjects?
  - ✧ Computationally unmanageable
  - ✧ Heterogeneity in data or experiment design across subjects
  - ✧ Model quality check at individual subject level

Toy example of group analysis: FSM

❑ Responses from a group of subjects under one condition

  ✧ What we have: $(\beta_1, \beta_2, \ldots, \beta_{10})$=(1.13, 0.87, …, 0.72)

❑ Centroid: average $(\beta_1+\beta_2+\ldots+\beta_{10})/10 = 0.92$ is not enough

  ✧ Variation/reliability measure: diversity, spread, deviation

❑ Model building

  ✧ Subject i's response = group average + deviation of subject i: simple model GLM (one-sample t-test)

$$\hat{\beta}_i = b + \epsilon_i, \, \epsilon_i \sim N(0, \sigma^2)$$

  ✧ If individual responses are consistent, ϵ_i should be small

   ▪ t-test: significance measure = $\dfrac{\hat{b}}{\hat{\sigma}/n}$

   ▪ 2 measures: b (dimensional) and t (dimensionless)

Models at Group Level: FSM

- Conventional approach: taking β (or linear combination of multiple β's) only for group analysis

    o Assumption: all subjects have same precision (reliability, standard error, confidence interval) about

    o All subjects are treated equally

    o Student's t-test: paired, one- and two-sample: not random-effects models in strict sense as usually claimed

    o AN(C)OVA, GLM, LME

- Alternative: taking both effect estimates and t-statistics

    o t-statistic contains precision information about effect estimates

    o Each subject is weighted based on precision of effect estimate

Classical ANOVA: 2 × 3 Mixed ANCOVA

✧ Factor A (Group): 2 levels (patient and control)

✧ Factor B (Condition): 3 levels (pos, neg, neu)

✧ Factor S (Subject): 15 ASD children and 15 healthy controls

✧ Covariate (Age): **cannot** be modeled; **no** correction for sphericity violation

$$F_{(a-1,a(n-1))}(A) = \frac{MSA}{MSS(A)},$$

$$F_{(b-1,a(b-1)(n-1))}(B) = \frac{MSB}{MSE},$$

$$F_{((a-1)(b-1),a(b-1)(n-1))}(AB) = \frac{MSAB}{MSE},$$

where

$$MSA = \frac{SSA}{a-1} = \frac{1}{a-1}\left(\frac{1}{bn}\sum_{j=1}^{a} Y_{.j.}^2 - \frac{1}{abn}Y_{...}^2\right),$$

$$MSB = \frac{SSB}{b-1} = \frac{1}{b-1}\left(\frac{1}{an}\sum_{k=1}^{b} Y_{..k}^2 - \frac{1}{abn}Y_{...}^2\right),$$

$$MSAB = \frac{SSAB}{(a-1)(b-1)} = \frac{1}{(a-1)(b-1)}\left(\frac{1}{n}\sum_{j=1}^{a}\sum_{k=1}^{b} Y_{.jk} - \frac{1}{bn}\sum_{j=1}^{a} Y_{.j.}^2 - \frac{1}{an}\sum_{k=1}^{b} Y_{..k}^2 + \frac{1}{abn}Y_{...}^2\right),$$

$$MSS(A) = \frac{SSS(A)}{a(n-1)} = \frac{1}{a(n-1)}\left(\frac{1}{b}\sum_{i=1}^{n}\sum_{j=1}^{a} Y_{ij.}^2 - \frac{1}{bn}\sum_{j=1}^{a} Y_{.j.}^2\right),$$

$$MSE = \frac{1}{a(b-1)(n-1)}\left(\sum_{i=1}^{n}\sum_{j=1}^{a}\sum_{k=1}^{b} Y_{ijk}^2 - \frac{1}{n}\sum_{j=1}^{a}\sum_{k=1}^{b} Y_{.jk} - \frac{1}{b}\sum_{i=1}^{n}\sum_{j=1}^{a} Y_{ij.}^2 + \frac{1}{bn}\sum_{j=1}^{a} Y_{.j.}^2 + \frac{1}{abn}Y_{...}^2\right)$$

Univariate GLM: 2 x 3 mixed ANOVA

- Group: 2 levels (patient and control)
- Condition: 3 levels (pos, neg, neu)
- Subject: 3 ASD children and 3 healthy controls

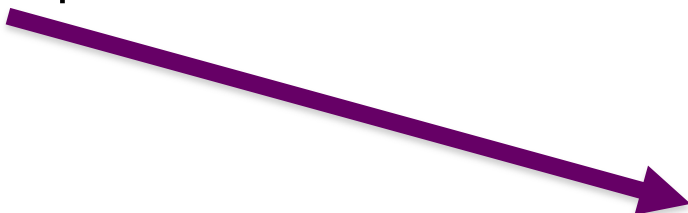*Difficult to incorporate covariates*
- *Broken orthogonality*

No correction for sphericity violation

$$
\mathbf{b} =
\begin{pmatrix}
\beta_{11} \\ \beta_{12} \\ \beta_{13} \\
\beta_{21} \\ \beta_{22} \\ \beta_{23} \\
\beta_{31} \\ \beta_{32} \\ \beta_{33} \\
\beta_{41} \\ \beta_{42} \\ \beta_{43} \\
\beta_{51} \\ \beta_{52} \\ \beta_{53} \\
\beta_{61} \\ \beta_{62} \\ \beta_{63}
\end{pmatrix}
$$

Subj	X_0	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
1	1	1	1	0	1	0	1	0	0	0
1	1	1	0	1	0	1	1	0	0	0
1	1	1	−1	−1	−1	−1	1	0	0	0
2	1	1	1	0	1	0	0	1	0	0
2	1	1	0	1	0	1	0	1	0	0
2	1	1	−1	−1	−1	−1	0	1	0	0
3	1	1	1	0	1	0	−1	−1	0	0
3	1	1	0	1	0	1	−1	−1	0	0
3	1	1	−1	−1	−1	−1	−1	−1	0	0
4	1	−1	1	0	−1	0	0	0	1	0
4	1	−1	0	1	0	−1	0	0	1	0
4	1	−1	−1	−1	1	1	0	0	1	0
5	1	−1	1	0	−1	0	0	0	0	1
5	1	−1	0	1	0	−1	0	0	0	1
5	1	−1	−1	−1	1	1	0	0	0	1
6	1	−1	1	0	−1	0	0	0	−1	−1
6	1	−1	0	1	0	−1	0	0	−1	−1
6	1	−1	−1	−1	1	1	0	0	−1	−1

$$
= \mathbf{X}
\begin{pmatrix}
\alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \\ \alpha_7 \\ \alpha_8 \\ \alpha_9
\end{pmatrix}
+ \mathbf{d}
\begin{pmatrix}
\delta_{11} \\ \delta_{12} \\ \delta_{13} \\
\delta_{21} \\ \delta_{22} \\ \delta_{23} \\
\delta_{31} \\ \delta_{32} \\ \delta_{33} \\
\delta_{41} \\ \delta_{42} \\ \delta_{43} \\
\delta_{51} \\ \delta_{52} \\ \delta_{53} \\
\delta_{61} \\ \delta_{62} \\ \delta_{63}
\end{pmatrix}
$$

32

Flexible Approach: Multivariate GLM

o **Group**: 2 levels (patient and control)

o **Condition**: 3 levels (pos, neg, neu)

o **Subject**: 3 ASD children and 3 healthy controls

o **Age**: quantitative covariate

$$B_{n \times m} = X_{n \times q} A_{q \times m} + D_{n \times m}$$

$$
\begin{array}{c}
Subj \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6
\end{array}
\begin{array}{ccc}
Pos & Neg & Neu \\
\left(\begin{array}{ccc}
\beta_{11} & \beta_{12} & \beta_{13} \\
\beta_{21} & \beta_{22} & \beta_{23} \\
\beta_{31} & \beta_{32} & \beta_{33} \\
\beta_{41} & \beta_{42} & \beta_{43} \\
\beta_{51} & \beta_{52} & \beta_{53} \\
\beta_{61} & \beta_{62} & \beta_{63}
\end{array}\right)
\end{array}
=
\begin{array}{ccc}
Int & Grp & Age \\
\left(\begin{array}{ccc}
1 & 1 & -6 \\
1 & 1 & 10 \\
1 & 1 & 4 \\
1 & -1 & -4 \\
1 & -1 & -1 \\
1 & -1 & -3
\end{array}\right)
\end{array}
\begin{array}{ccc}
Pos & Neg & Neu \\
\left(\begin{array}{ccc}
\alpha_{01} & \alpha_{02} & \alpha_{03} \\
\alpha_{11} & \alpha_{12} & \alpha_{13} \\
\alpha_{21} & \alpha_{22} & \alpha_{23}
\end{array}\right)
\end{array}
+
\begin{array}{ccc}
Pos & Neg & Neu \\
\left(\begin{array}{ccc}
\delta_{11} & \delta_{12} & \delta_{13} \\
\delta_{21} & \delta_{22} & \delta_{23} \\
\delta_{31} & \delta_{32} & \delta_{33} \\
\delta_{41} & \delta_{42} & \delta_{43} \\
\delta_{51} & \delta_{52} & \delta_{53} \\
\delta_{61} & \delta_{62} & \delta_{63}
\end{array}\right)
\end{array}
\begin{array}{c}
Subj \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6
\end{array}
$$

33

Inter-Subject Correlation (ISC)

- Analysis methodology
  - Regression with task-related regressors won't work
  - Voxel-wise correlation between any subject pair
    - $n = 4$ subjects => 6 ISC; $n = 5$ subjects $\rightarrow$ 10 ISC
    - n subjects => $n(n-1)/2$ ISC – which are not all independent!
  - How to go about group analysis?
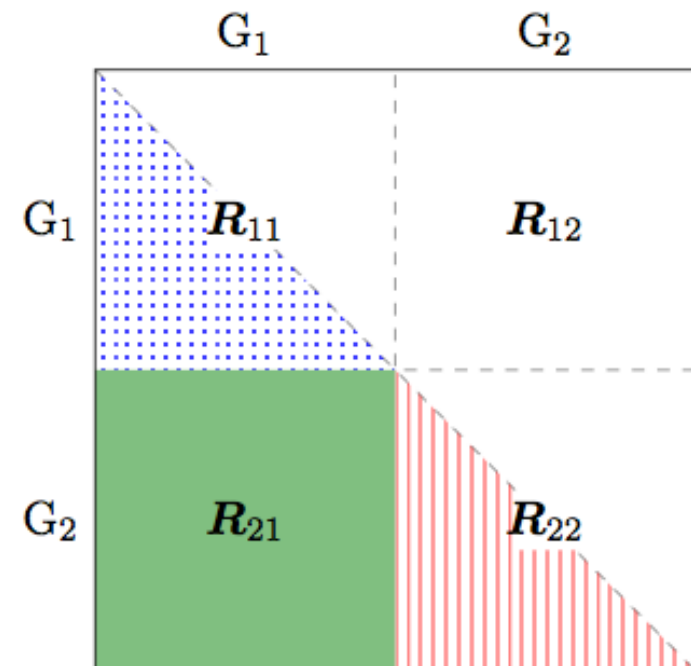
Inter-Subject Correlation (ISC)

- Analysis methodology
  - How to go about group analysis?
    - Difficulty: The ISCs are not independent with each other
    - The correlations are correlated themselves!
  - Solutions
    - Permutations
    - Bootstrapping
    - LME

	Z_{21}	Z_{31}	Z_{41}	Z_{51}	Z_{32}	Z_{42}	Z_{52}	Z_{43}	Z_{53}	Z_{54}
Z_{21}	1	ρ	ρ	ρ	ρ	ρ	ρ	0	0	0
Z_{31}	ρ	1	ρ	ρ	ρ	0	0	ρ	ρ	0
Z_{41}	ρ	ρ	1	ρ	0	ρ	0	ρ	0	ρ
Z_{51}	ρ	ρ	ρ	1	0	0	ρ	0	ρ	ρ
Z_{32}	ρ	ρ	0	0	1	ρ	ρ	ρ	ρ	0
Z_{42}	ρ	0	ρ	0	ρ	1	ρ	ρ	0	ρ
Z_{52}	ρ	0	0	ρ	ρ	ρ	1	0	ρ	ρ
Z_{43}	0	ρ	ρ	0	ρ	ρ	0	1	ρ	ρ
Z_{53}	0	ρ	0	ρ	ρ	0	ρ	ρ	1	ρ
Z_{54}	0	0	ρ	ρ	0	ρ	ρ	ρ	ρ	1

Multiple Testing Correction

- **Two types of errors**
  - *What is H_0 in FMRI studies? H_0 = no effect (activation, difference, …) at a voxel*

  - Type I error = Prob(reject H_0 when H_0 is true) = false positive = p value
    - Type II error = Prob(accept H_0 when H_1 is true) = false negative = β
    - **power** = $1-\beta$ = probability of detecting true activation
  - Goal: control type I error rate while increasing power (decreasing type II errors)
  - Significance level **α** (magic number 0.05) : $p < \alpha$

Justice System: Trial

Hidden Truth

	Defendant Innocent	Defendant Guilty
Reject Presumption of Innocence (Guilty Verdict)	**Type I Error** (defendant very unhappy)	**Correct**
Fail to Reject Presumption of Innocence (Not Guilty Verdict)	**Correct**	**Type II Error** (defendant very happy)

Statistics: Hypothesis Test

Hidden Truth

	H_0 True Not Activated	H_0 False Activated
Reject H_0 (decide voxel is activated)	**Type I Error** (false positive)	**Correct**
Don't Reject H_0 (decide voxel isn't activated)	**Correct**	**Type II Error** (false negative)

Multiple Testing Correction
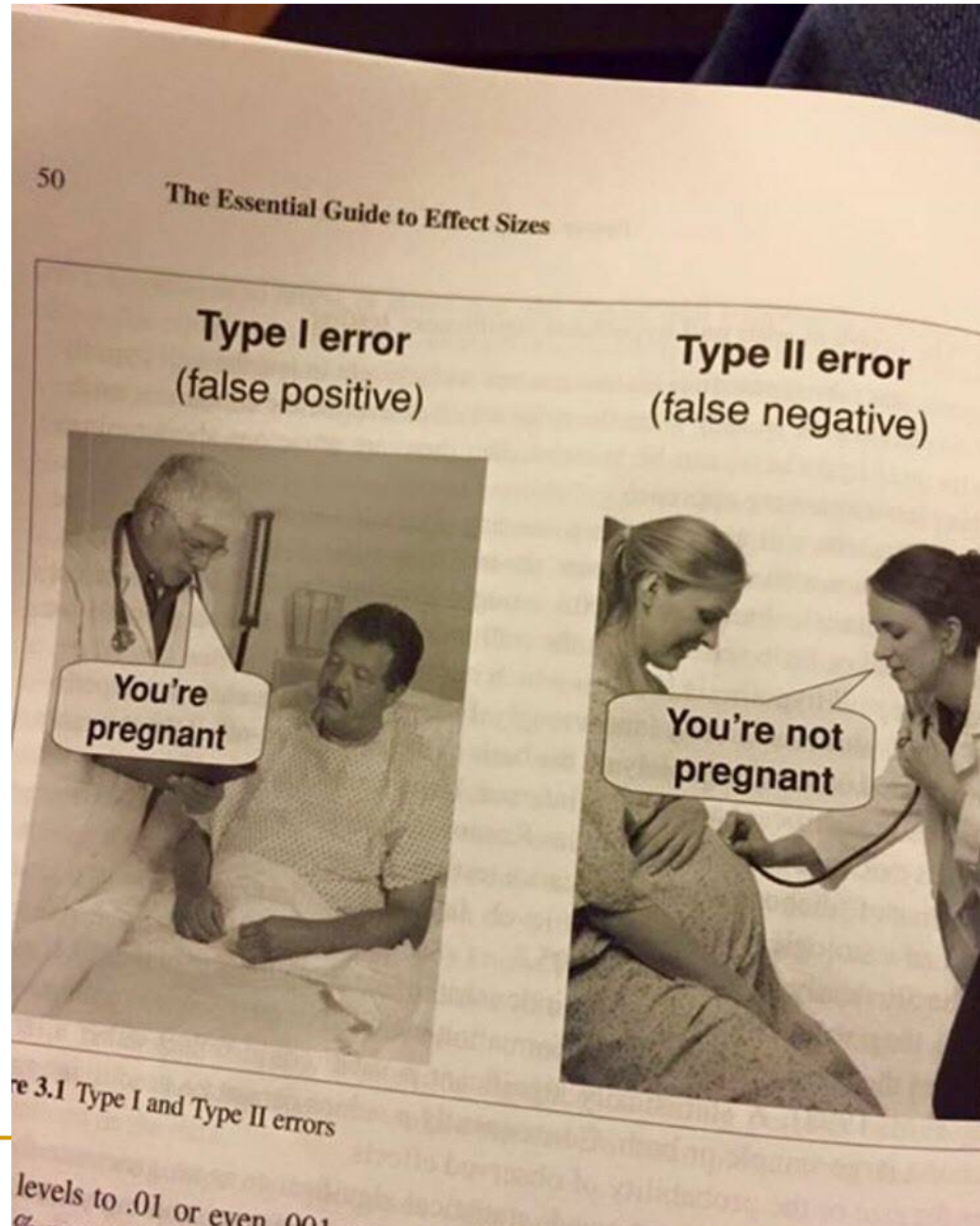
Multiple Testing Correction

- **Two types of correction**
  - Same test repeated many times (number of voxels)
    - Family-wise error (FWE)
    - Counterbalance between cluster size and voxel-wise significance
    - Difficulties: do not want to over-penalize due to spatial correlation, but spatial structure is neither Gaussian nor homogeneous!
    - Correction via permutations with quantifying the counterbalance
  - Multiple tests in a study: six pairwise comparisons among four levels of a factor
    - Little attention in neuroimaging community

Miscellaneous

- Science is about reproducibility
  - Widespread obsession with p-values in FMRI
    - Colored blobs of t-values
    - Peak voxel selected based on peak t-value
  - **Unacceptable** in some fields if only p-value is reported
    - Basic and Applied Social Psychology bans p-value
    - Neuroimaging: an exception currently!
- 2-tier approach
  - Start with a liberal thresholding of p: 0.05 or even 0.1
  - Report rigorous results (multiple testing correction)
  - Also report the clusters under voxel-wise p of 0.05 but still have some spatial extent with some cautionary words

Problems of conventional statistics

- ❑ *p*-value problem (obsession) under NHST
  - ✧ Strawman: null hypothesis H_0 (e.g., no activation)
  - ✧ $p = \text{probability}(\text{data} \mid H_0)$
    - ✧ Extent or evidence of discrepancy of model with H_0
    - ✧ **Not** updated probability about the truth of H_0
    - ✧ Disconnected with the investigator: usually **not** a measure of interest
- ❑ Damaging effects of *p*-value obsession
  - ✧ Dichotomization: statistically significant vs insignificant
  - ✧ Vulnerable to misinterpretation; activated vs inactivated
    - ✧ Difference between a statistically significant result and an insignificant one is **not** necessarily statistically significant
 - ✧ *p*-hacking (data massage, small volume correction, etc.)

Problems of conventional statistics

- Many branching points in FMRI data processing
    - Different smoothing: 4-10mm
    - Different calibration/scaling strategy: voxel-wise, whole brain
    - Handling head motion effect
- Segmented modeling
    - Splitting into individual and group levels
    - Voxel-wise modeling: multiple testing problem

Miscellaneous

❑ Correlation vs. causation

✦ Incidence of childhood leukemia: several times greater in certain Denver suburbs than in other parts of the country

▪ High-voltage power lines?

✦ Lots of correlation analysis in FMRI
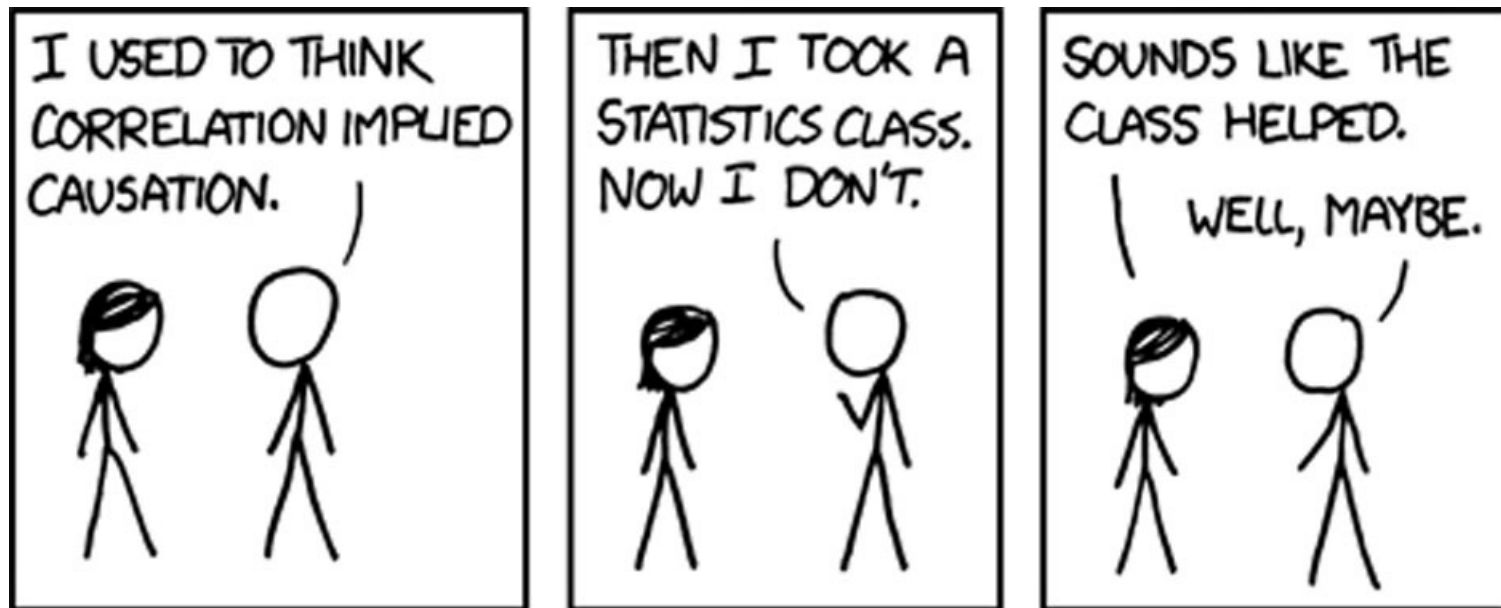
Lastly

- Essentially all models are wrong, but some are useful. (George E. P. Box)

- Statistics are like a bikini. What they reveal is suggestive, but what they conceal is vital. (Aaron Levenstein)