Slides

Day 1, Session 1

Introduction

Workshop 7-8th July 2014
Ghent University

Goals

- Conduct SPM analyses
- Interpret SPM's results
- Report SPM methods and results

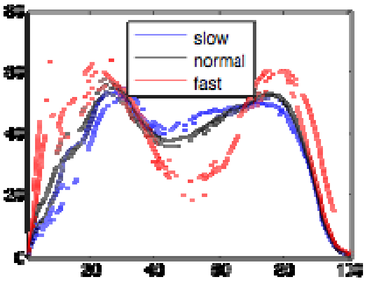
Workshop 7-8th July 2014
Ghent University

Not-Goals

- Become highly proficient in Matlab
- Understand all aspects of SPM
- Understand all aspects of conventional statistics
- Focus on structuring and inputting data

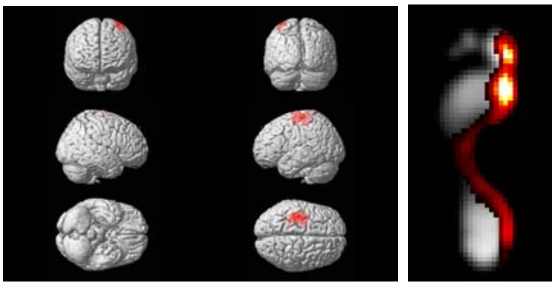
Workshop 7-8th July 2014
Ghent University

Dataset



SPM

Statistical Parametric Mapping



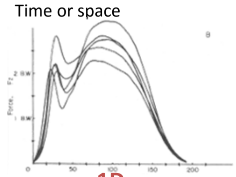
Suitable for what biomechanical data?

a) Spatiotemporally smooth

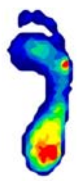
- Sampled above Nyquist frequency
- Biological tissue viscoelasticity

b) Bounded

- Time or space



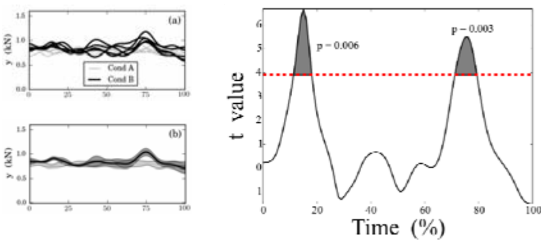
1D



2D

SPM

Statistical Parametric Mapping

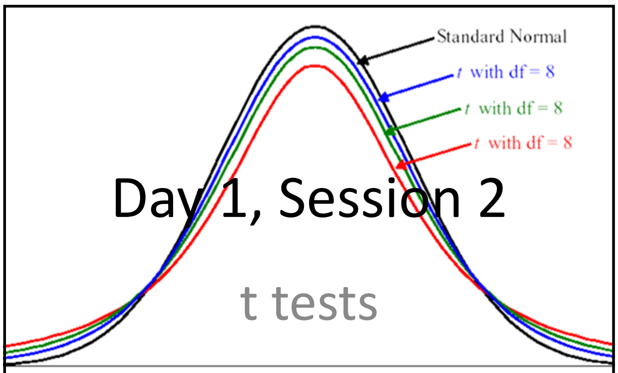


Questions

- Have you heard of Matlab?
- Have you processed data in Matlab?
- Have you conducted statistical tests in Matlab?
- Have you validated probability computations in Matlab?

MATLAB

- “MATrix LABoratory”
- Numerical array computations
- Useful for data analysis
- Broad functionality:
  - Statistics, optimization, signal processing, dynamic simulation, etc.



Day 1, Session 2

t tests

Goals

- t-test recap
- Univariate vs 1DSPM calculations
- Run t-tests in Matlab
- Interpret output variables

T-tests

- one-sample t-test
  - Used to test the difference between a single dataset and a constant (usually zero)
- paired t-test
  - Used to test for a differences in repeated measurements of a sample
- two-sample t-test
  - Used to test for a difference between two samples

Experimental scenarios

1. One subject, multiple trials, two tasks
2. 100 subjects, one body mass value each, compared against national average
3. Surgical Procedure A vs. Surgical Procedure B on 40 subjects per procedure
4. 100 subjects, tested at ages 20 and 30
5. one subject, measured multiple times, compared to a "normal" database
6. Controls vs. Patients
7. Multiple subjects performing two different tasks once each

First, we need to set up our null hypothesis.

Null Hypothesis (H_0)

– Predicts no significant difference

e.g. There is no significant difference in vertical ground reaction forces in injured versus uninjured runners

Univariate
two-sample t-test calculation

$$t = \frac{\bar{y}_B - \bar{y}_A}{\sqrt{\frac{1}{J}(s_A^2 + s_B^2)}}$$

Univariate
two-sample t-test

- 1. Compute mean values y_A and y_B
- 2. Compute the standard deviations s_A and s_B
- 3. Compute the t test statistic
- 4. Conduct statistical inference. Use α and the t distribution to compute the critical threshold. If $t > t_{critical}$ reject null hypothesis
- 5. Compute exact p-value using t and the univariate t-distribution.

Univariate
ex_ttest_twosample.m

```
%(1) Conduct 0D test using spm1d:
spm  = spm1d.stats.ttest2(yA, yB);
spmi = spm.inference(0.05, 'two_tailed', false);
```

yA = 10 x 1 double

yB = 10 x 1 double

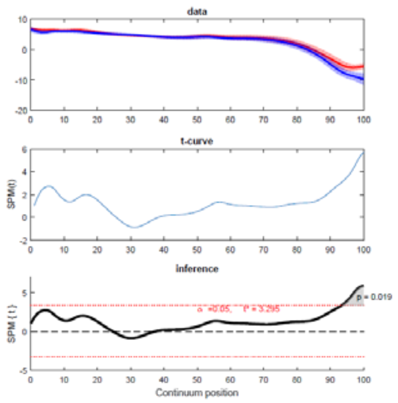
Univariate
reporting

- For example:

There was no significant difference between the male and female reaction times, $t(9) = 0.52$, $p = 0.15$.

SPM1D
two-sample *t*-test

1. Compute mean fields $y_A(q)$ and $y_B(q)$
2. Compute the st. dev. of fields $s_A(q)$ and $s_B(q)$
3. Compute the *t* test statistic field
4. Conduct statistical inference.
Use α and the RFT *t* distribution to compute the critical threshold. If SPM{*t*} > *t* critical reject null hypothesis for suprathreshold clusters.
5. Compute exact *p*-value for each cluster using cluster size and RFT distribution(s) for SPM{t} topology.



SPM1D
calculation comparison

$$t = \frac{\bar{y}_B - \bar{y}_A}{\sqrt{\frac{1}{J}\left({s_A}^2 + {s_B}^2\right)}}$$

$$SPM\{t\} \equiv t(q) = \frac{\bar{y}_B(q) - \bar{y}_A(q)}{\sqrt{\frac{1}{J}\left({s_A}^2(q) + {s_B}^2(q)\right)}}$$

The calculation of *t*(*q*) does not mean that we are "doing 'q' t-tests".

SPM1D
t-test Matlab function names

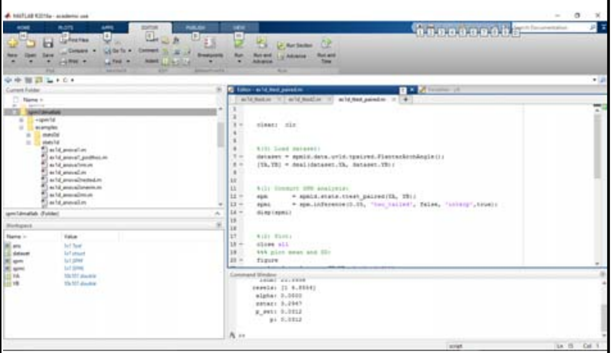
spm1d.stats.ttest(…)	One-sample t test
spm1d.stats.ttest_paired(…)	Paired t test
spm1d.stats_ttest2(….)	Two-sample test

SPM1D
ex1d_ttest2.m

```
%(1) Conduct SPM analysis:
spm  = spm1d.stats.ttest2(YA, YB);
spmi = spm.inference(0.05,'two_tailed',true);
```

yA = 10 x 101 double
yB = 10 x 101 double

Over to Matlab



Questions resulting from the analysis of these datasets?

Helpful tips?

Day 1, Session 3

Probability and Random Field Theory

spm1D

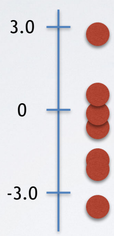
Goals

- Describe the meaning of *p* values for 1D analyses
- Simulate experiments using smooth, random 1D data to validate *p* values

spm1D

One can’t use one’s knowledge about randomness in oranges to make probabilistic conclusions regarding apples.

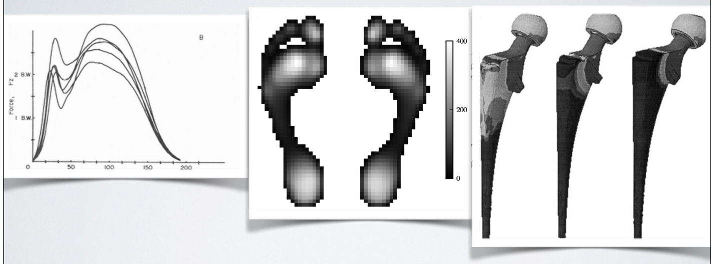
Zero-dimensional data



Body mass
Body height
Jump height

Variables which do not change in space or time

n-dimensional data



One can't use one's knowledge about randomness in **0D data** to make probabilistic conclusions regarding ***n*D data**.

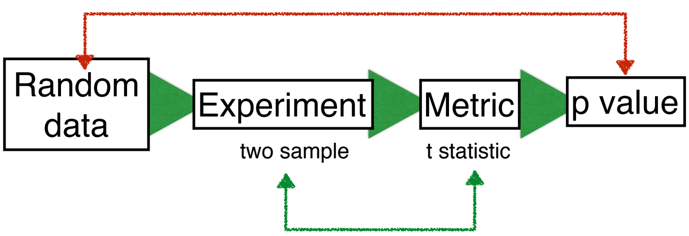
Goals

- Describe the meaning of *p* values for 1D analyses
- Simulate experiments using smooth, random 1D data to validate *p* values

What is a *p* value?

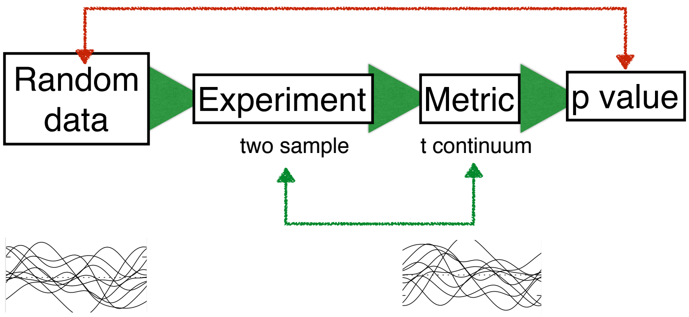
Demo

Summary



Demo 1D

Summary 1D



What is a *p* value?

The probability that a completely random *n*D process will yield a particular result.

Goals

- Describe the meaning of *p* values for 1D analyses

- Simulate experiments using smooth, random 1D data to validate *p* values

Summary

t values describe experimental data

p values describe random data

Use an ***nD*** model of randomness to make probabilistic conclusions regarding ***nD* data**

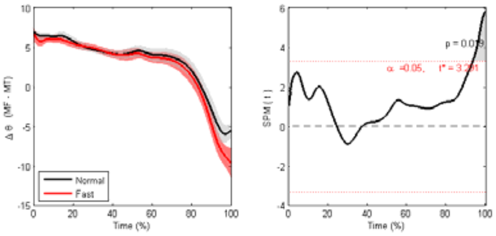
p is not:

x The probability that the null hypothesis is true

x The probability that the alternative hypothesis is false

x The probability that the observed result is random

✅ $\mathrm{P}(\mathrm{data}|H_0)$



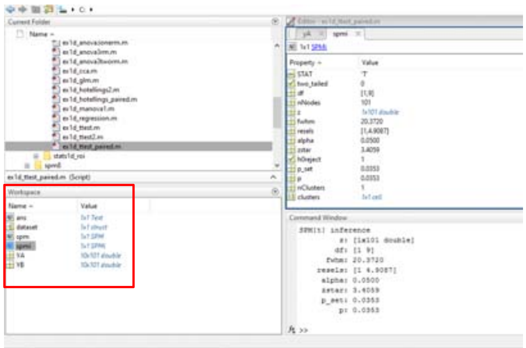
Day 1 Session 4
Interpretation of t-tests

“Is it significant?”

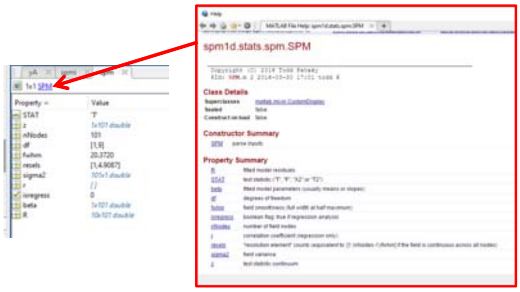
Plan

1. Interpret t-test results
2. Describe the methods used
3. Present the results
4. Reviewer hints and tips

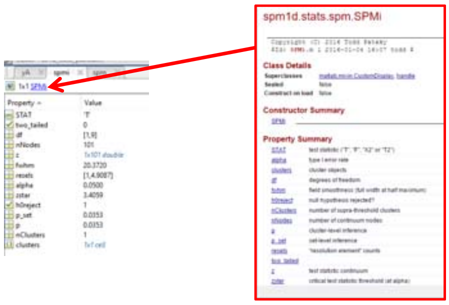
1. Interpretation



the “SPM” variable

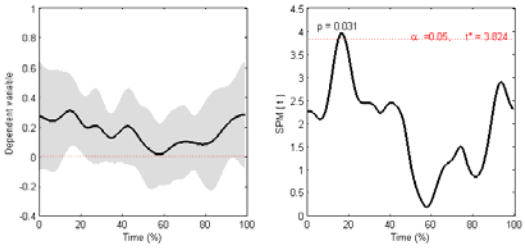


the “SPMi” variable



Interpretation

ex1d_ttest.m

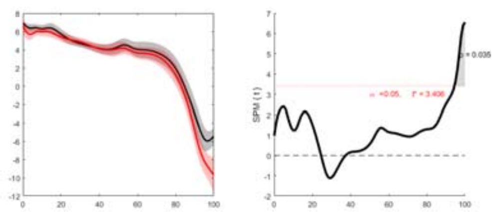


>> disp(spmi)

SPM{t} inference
z: [1x100 double]
df: [1 9]
fwhm: 11.2615
resels: [1 8.7910]
alpha: 0.0500
zstar: 3.8213
p_set: 0.0310
p: 0.0310

Key aspects – Interpretation

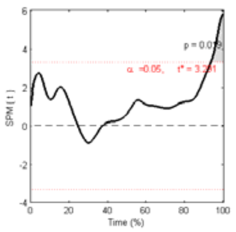
ex1d_ttest_paired.m



>> disp(spmi)

SPM{t} inference
z: [1x101 double]
df: [1 9]
fwhm: 20.3720
resels: [1 4.9087]
alpha: 0.0500
zstar: 3.4059
p_set: 0.0353
p: 0.0353

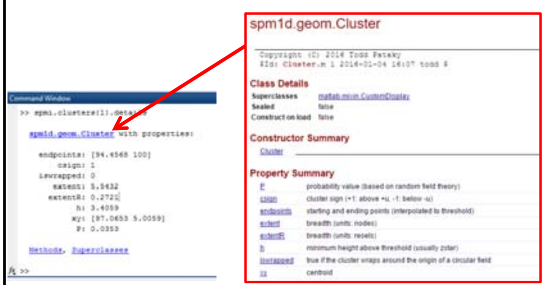
Cluster specific p-values



Identically smooth random 1D data would produce a cluster of this breadth with a probability of p=0.019

Clusters

`spmi.clusters{1}.details`



2. Methods

What are the contents of a good methods section?

SPM Methods

a) Statistical tests used
b) SPM code & analysis software
c) Refer to key SPM/RFT literature
d) Define terminology
e) Specify alpha – correction?
f) How results will be interpreted

[also relevant: data treatment, smoothing, averaging]

a) Statistical tests used

- e.g. SPM One sample *t*-test
  - One or two-tailed?
  - Dependent variable tested
  - Independent variable

"A SPM two-tailed two-sample t-test was used to compare male versus female knee angles."

b) SPM code

- Can refer to http://www.spm1d.org/ for open source code
- The current Python version of **spm1d** is:
- The current Matlab version of **spm1d** is:

Software

Python 2.7.2; Enthought Python Distribution, Austin, TX.

Matlab R2016a (8.3.0.532), The Mathworks Inc, Natick, MA.

[Home → Help → About Matlab]

c) Key SPM/RFT literature

SPM literature (Neuroimaging)

- Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE, Penny WD. (Eds.) Statistical parametric mapping: the analysis of functional brain images. London: Elsevier; 2007.
  - A book that describes the SPM analysis concepts with specific application to brain images.
- Friston KJ, Holmes AP, Worsley KJ, Poline JB, Frith CD, Frackowiak RSJ (1995). Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping* **2**, 189–210.
- SPM documentation repository, Wellcome Trust Centre for Neuroimaging: http://www.fil.ion.ucl.ac.uk/spm/doc/

SPM/RFT Literature (Biomechanics)

- Pataky, T. C. . (2012). One-dimensional statistical parametric mapping in python. *Computer Methods in Biomechanics and Biomedical Engineering,* 15(3):295–301.
- Pataky, T. C. (2016). Rft1d: Smooth one-dimensional random field upcrossing probabilities in python. *Journal of Statistical Software*, page in press.

d) Terminology –

- SPM vs SPM{*t*}

SPM refers to the overall methodological approach

SPM{t} to the scalar trajectory variable

- Suprathreshold cluster

"adjacent points of the SPM{t} curve often exceed the critical threshold, we therefore call these "supra-threshold clusters".

- Critical threshold

"Value at which only α % (5%) of smooth random curves would be expected to cross"

e) Specify alpha

- **You may wish to correct alpha for multiple comparisons / dependent variables.**
- **Remember this would have to be entered manually in the code.**

$$p_{critical} = 1 - (1 - \alpha)^{1/N}$$

"To retain a family-wise Type 1 error rate of α = 0.05 we adopted a Šidák corrected threshold of 0.012 for four comparisons."

spm 1D

f) How results will be interpreted

- What does a high value of SPM{*t*} mean?
- What is the consequence of crossing the critical threshold (or not)?

"if the SPM{t} trajectory crosses the critical threshold at any time node, the null hypothesis is rejected"

spm 1D

An example method

- Refer to handout
  "Day1Session4 - Example method write-up.docx"
- Task – Using the colour-coding in the bullet-points, colour the text in the example write-up to match the relevant bullet-point.
  The first sentence has been started as an example.
  What could be more concise?

spm 1D

Example methods to refer to

- Pataky TC (2010). Generalized n-dimensional biomechanical field analysis using statistical parametric mapping. *Journal of Biomechanics* 43, 1976-1982.
- Pataky TC (2012) One-dimensional statistical parametric mapping in Python. *Computer Methods in Biomechanics and Biomedical Engineering*. 15, 295-301.
- Pataky TC, Robinson MA, Vanrenterghem J (2013). Vector field statistical analysis of kinematic and force trajectories. *Journal of Biomechanics* 46 (14): 2394-2401.

Applications

- Vanrenterghem, J., Venables, E., Pataky, T., Robinson, M. (2012). The effect of running speed on knee mechanical loading in females during side cutting. *Journal of Biomechanics*, *45*, 2444-2449.
- De Ridder, R., Willems, T., Vanrenterghem, J., Robinson, M., Pataky, T., Roosen, P. (2013). Gait kinematics of subjects with chronic ankle instability using a multi-segmented foot model. *Medicine and Science in Sports and Exercise*, 45, 2129-2136.
- Robinson, M.A., Donnelly, C.J., Tsao, J., Vanrenterghem, J. (2014). Impact of knee modelling approach on indicators and classification of ACL injury risk. Medicine & Science in Sports & Exercise, 46 (7), 1269-1276.

3. Presentation of results

- Key information to present:

a) Was the critical threshold exceeded?
b) Direction of effect
c) Consequence for the null hypothesis
d) Descriptive data:
    critical threshold, p-value/s, number of supra-threshold clusters, extent of clusters, degrees of freedom.

Figure caption

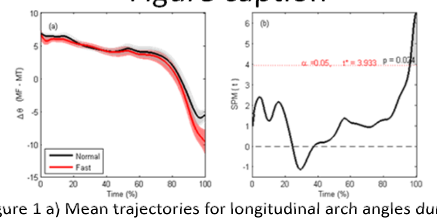


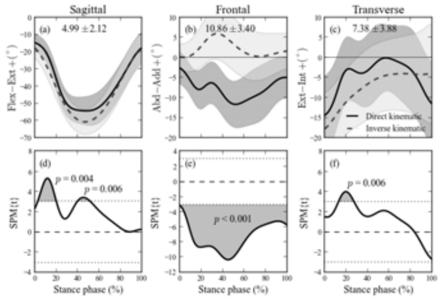
Figure 1 a) Mean trajectories for longitudinal arch angles during normal (black) and fast (red) walking. b) The paired samples t-test statistic SPM {*t*}. The critical threshold of 3.933 (red dashed line) was exceeded at time = 96% with a supra-threshold cluster probability of p=0.024 indicating a significantly more negative angle in the fast condition.

Results section example

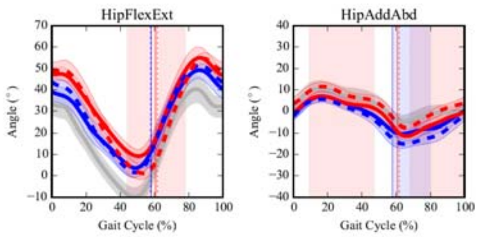
Mean longitudinal arch angles during normal and fast walking were highly similar for the majority of time (figure 1a). However one supra-threshold cluster (96-100%) exceeded the critical threshold of 3.933 as the ΔΘ(MF-MT) angle during fast walking was significantly more negative[1] than during normal walking (figure 1b). The precise probability that a supra-threshold cluster of this size would be observed in repeated random samplings was p=0.024. The null hypothesis was therefore rejected.

~~SPM{t}$_{max}$ = 6.55, p=0.024, d.f.=1,9~~

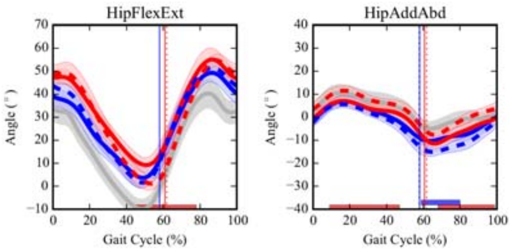
Example 1: Data and stats



Example 2: Shading



Example 3: Small bars



4. Reviewers – Hints and tips

- Anticipate some resistance
- Be consistent throughout
- Present statistical results with original data
- Highlight similarity to univariate interpretation
- Refer to papers that use similar analyses
- Benefits of SPM vs PCA, FDA etc
- Use supplementary material where appropriate

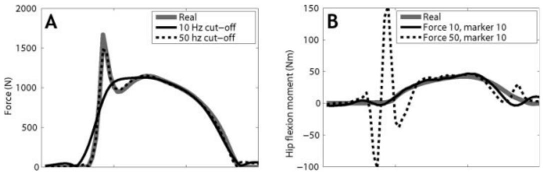
Tasks – Analyse and report…

Analyse and report the results of the data
paired_data.mat

1. Write the methods for the scenario below
   - Knee flexion angle data during running
   - 34 male participants
   - Two speed conditions
   - Data are normalised to the stance phase
2. Write a figure caption
3. Write a results section

Day 2, Session 1

Registration and smoothing



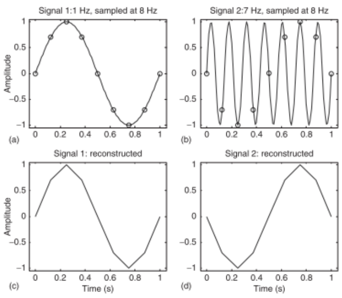
Goals

- Refresh smoothing and registration principles of biomechanical data
- Identify how smoothness and registration can affect SPM outcome

Overview

- Sampling of biomechanical signals
- Smoothing of biomechanical signals
- Registration of biomechanical signals
- Exercise on effect smoothing on SPM

Sampling biomechanical data

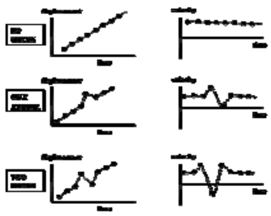


Smoothing biomechanical data

Remove artefact ⟺ Retain signal

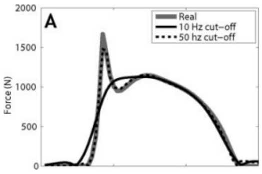
Smoothing

Kinematic data



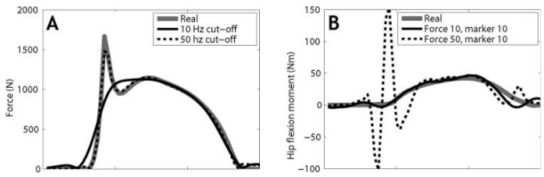
Displacement: $x(t) = \sin t$; Noise $(t) = 0.001\sin 50t$; Signal to noise ratio: 1:0.001.
Velocity: $v(t) = \cos t$; Noise $(t) = 0.05\cos 50t$; Signal to noise ratio: 1:0.05.
Acceleration: $a(t) = -\sin t$; Noise $(t) = -2.5\sin 50t$; Signal to noise ratio: 1:2.5.

Smoothing
Force data



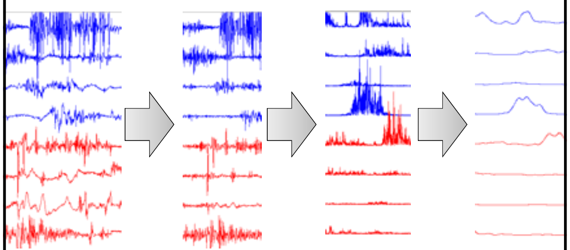
Smoothing
Kinetic data

GRF **Joint moment**

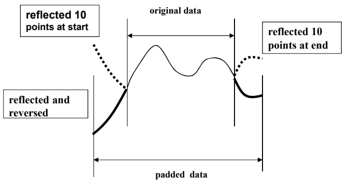


Kristianslund et al. (2012) Effect of low-pass filtering on joint moments from inverse dynamics: implications for injury prevention. J Biomech, 45, 666-671.

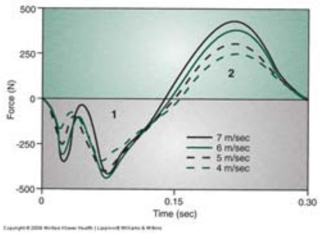
Smoothing
EMG data



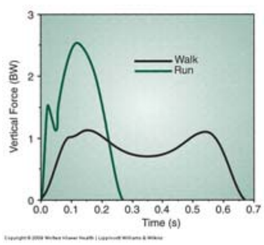
End point effects



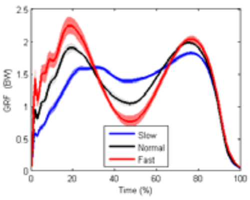
Registration



Registration



Registration



Registration

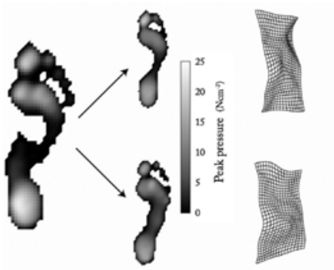
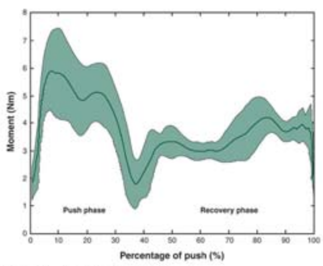


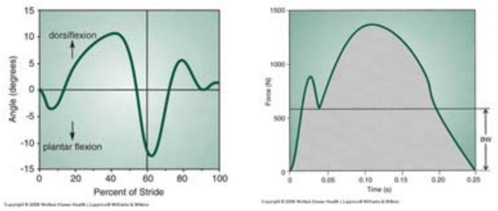
Image taken from Pataky et al., Gait Posture, Vol. 29, pp. 477-482, 2009

Registration



Linear registration

- Temporal context is retained uniformly



Relevance to SPM

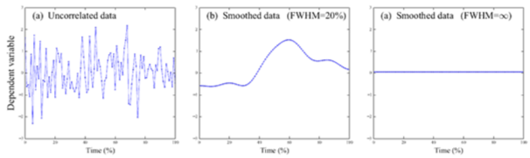
- Sampling affects SPM when Nyquist is violated
- Smoothing affects dependence 'nodes'

$$x_i' = a_0 x_i + a_1 x_{i-1} + a_2 x_{i-2} - b_1 x_{i-1}' - b_2 x_{i-2}'$$

=> Affects SPMi (see next slide)

Inference

RFT handles the issue of multiple comparisons and statistical inference . (Adler, 1981)



If neighbouring data are completely uncorrelated we'd conduct 100 independent t tests. If data are correlated then we have <100 independent tests.

The temporal (or spatial) gradient estimates # independent processes

Relevance to SPM

- Sampling affects SPM when Nyquist is violated
- Smoothing affects dependence ‘nodes’

$$x'_i = a_0 x_i + a_1 x_{i-1} + a_2 x_{i-2} - b_1 x'_{i-1} - b_2 x'_{i-2}$$

=> Affects SPMi

- Registration affects t-curve

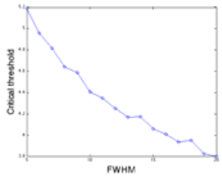
=> Affects SPM{t}

Exercise

Smoothness of group trajectories affects outcome SPM by lowering threshold

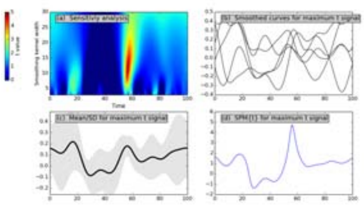
– Can we investigate relationship smoothness and threshold?

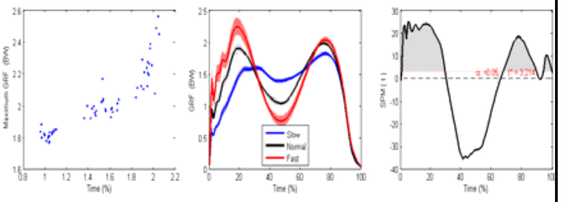
```
z = randn1d(J, Q, fwhm);
spm = spm1d.stats.ttest(z);
spmi = spm.inference(0.05);
```



Sensitivity of data processing

e.g. treat smoothness as a sensitivity measure





Day 2 Session 3

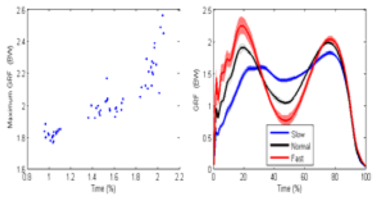
Interpretation of Regression & ANOVA

Plan

- Regression example
  - Code
  - Interpretation
- ANOVA example
  - Code
  - Interpretation

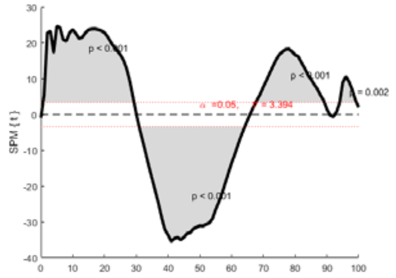
Regression

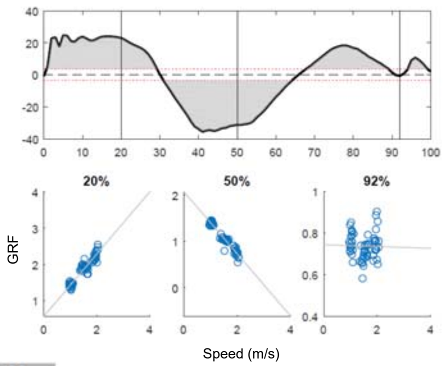
```
%(1) Conduct SPM analysis:
spm  = spm1d.stats.regress(Y, x);
spmi = spm.inference(0.05, 'two_tailed', true);
```



"x" independent variable
walking speed
"Y" dependent variable
vertical ground
reaction force

SPM output





Why a SPM{t} value?

The t value is the statistic upon which statistical inferences (p values) are based.

The regression coefficient r and t value are alternative expressions of effect magnitude, and map directly to each other given the sample size.

spm.r is the correlation coefficient

Regression interpretation

There was a significant relationship between walking speed and vGRF. A greater walking speed significantly increased the vGRF during the first and last 30% stance but significantly reduced GRF from ~30-70% stance. As random data would produce this effect <5% time. The null hypothesis was therefore rejected.

Figure caption

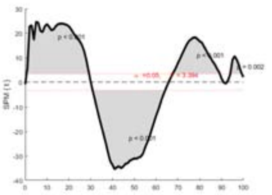
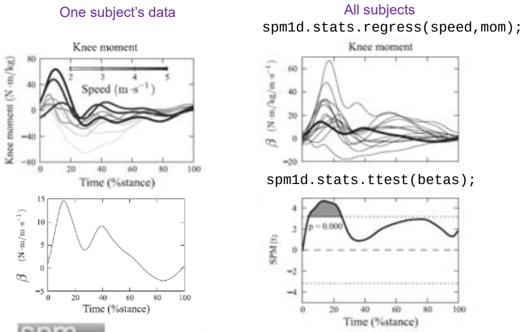


Figure 1. A statistical inference curve indicating significant relationships between walking speed and vertical ground reaction force. There were significant positive relationships during early and late stance and a significant negative relationship during mid-stance.

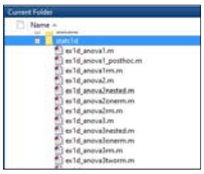
Random effects analysis



Vanrenterghem et al. (2012) JoB

ANOVA

- Many ANOVA types



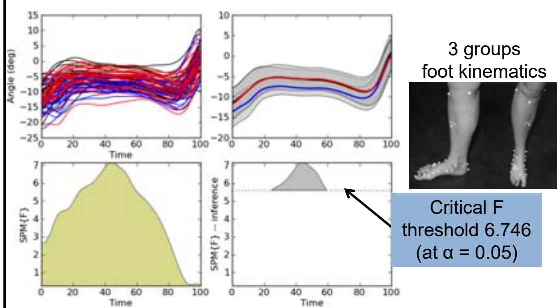
SPM code

```
%(1) Conduct SPM analysis:
spm        = spm1d.stats.anova1(Y, A);
spmi       = spm.inference(0.05);
```

```
%(1) Conduct SPM analysis:
spm_bs     = spm1d.stats.anova1(Y, A);
      %between-subjects model
spm        = spm1d.stats.anova1rm(Y, A, SUBJ);
      %within-subjects model
spmi       = spm.inference(0.05);
```

One-way ANOVA

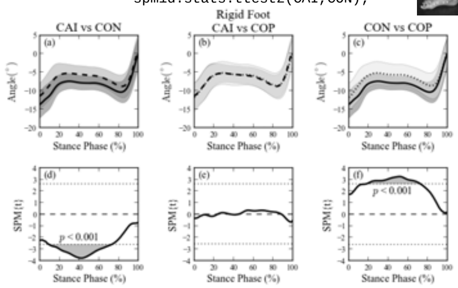
`spm1d.stats.anova1(CAI,COP,CON);`



De Ridder et al. (2013). MSSE

ANOVA post-hoc

```
spm1d.stats.ttest2(CAI,CON);
```



De Ridder et al. (2013). MSSE

AVONA interpretation

- There was a significant difference in angle between the three groups between ~20-60% stance. Identically smooth random 1D data would produce a cluster of this breadth with a probability of $p<0.05$. Post-hoc independent *t*-tests showed that CAI and COP were both significantly greater than the CON group.

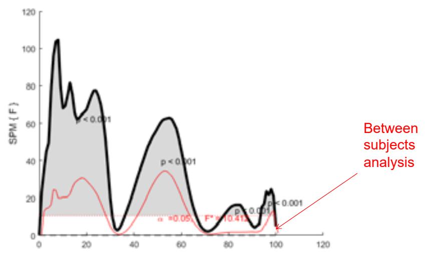
One way RM ANOVA

```
%(1) Conduct SPM analysis:
spm_bs     = spm1d.stats.anova1(Y, A);
     %between-subjects model
spm        = spm1d.stats.anova1rm(Y, A, SUBJ);
     %within-subjects model
spmi       = spm.inference(0.05);
```

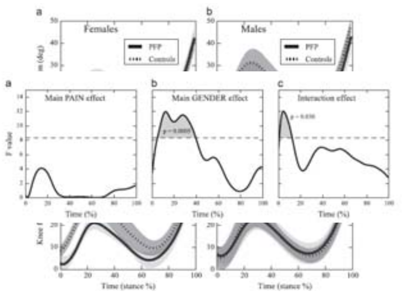
```
Warning: Only one observation per subject found.
Residuals and inference will be approximate. To
avoid approximate residuals: (a) Add multiple
observations per subject and per condition, and (b)
ensure that all subjects and conditions have the
same number of observations.

> In
spm1d.stats.anova.designs.ANOVA1rm/check_for_single_
responses (line 31)
  In spm1d.stats.anova1rm (line 9)
```

SPM output



Two-way ANOVA



Pataky et al. (2015). JoB

Tasks – Analyse and report…

- Finish ANOVA worksheet
- Analyse the example datasets 1-4
  1. Write the analysis code
  2. Plot the results
  3. Write a figure caption
  4. Write up the results

spm 1D

Day 2, Session 4

Introduction to Vector Analysis with SPM

Topics

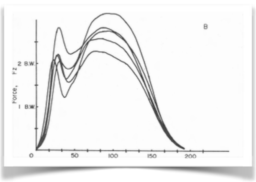
- Overview
- False positives
- Vector field statistics
- Reminder of SPM limitations
- Concluding remarks

Topics

- Overview
- False positives
- Vector field statistics
- Reminder of SPM limitations
- Concluding remarks

Scalar 1D

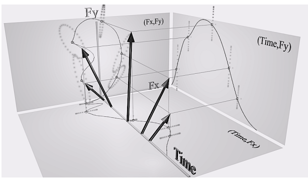
Univariate 1D



t tests
regression
ANOVA

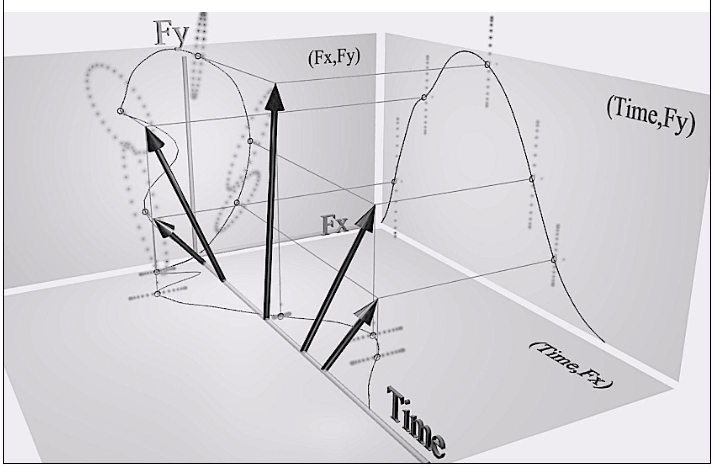
Vector 1D

Multivariate 1D

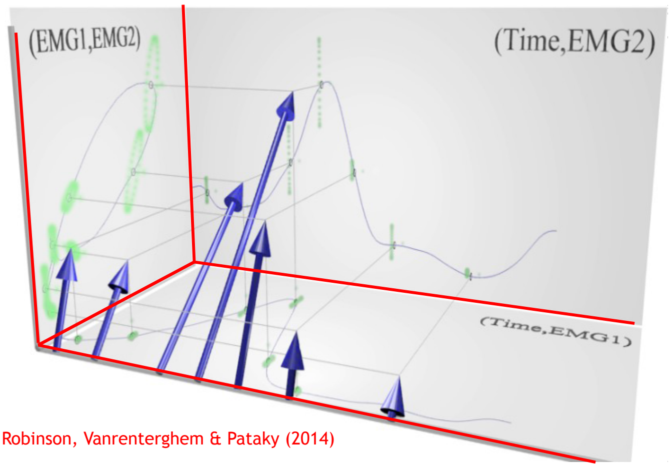


Hotelling's T^2 tests
CCA
MANOVA

Vector data: forces



Vector data: EMG



Robinson, Vanrenterghem & Pataky (2014)

Topics

- Overview
- False positives
- Vector field statistics
- Reminder of SPM limitations
- Concluding remarks

False positives

Demo

Methods

- Assembled ~1000 vector trajectories from public datasets
- Estimated the median smoothness for different classes of data
  - Kinematics, dynamics, EMG
- Computed false positive rates for 0D thresholds
  - Analytically
  - Validated using simulation

Results

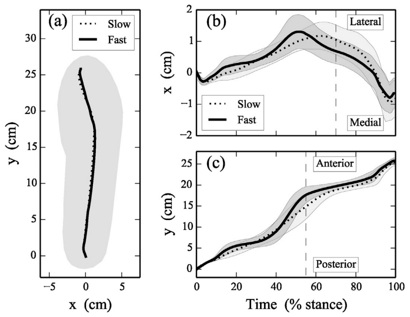
- Probability of a false positive:
  - One scalar trajectory
    - $p = 0.382$
  - One 3-component vector trajectory
    - $p = 0.764$
  - Two 3-component vector trajectories
    - $p = 0.945$

Topics

- Overview
- False positives
- Vector field statistics
- Reminder of SPM limitations
- Concluding remarks

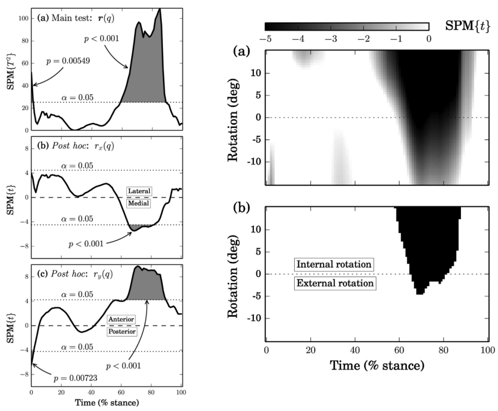
spm 1D

Example 1: COP

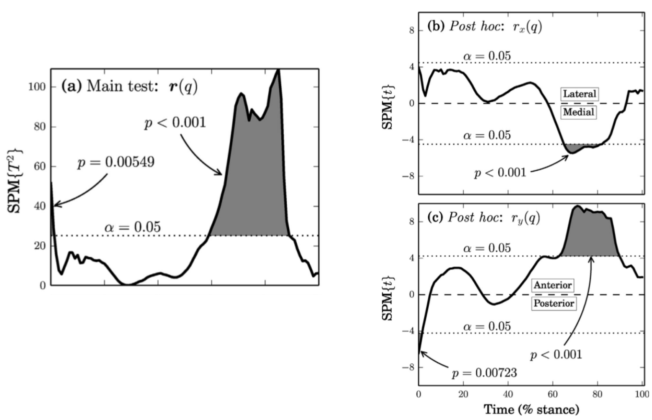


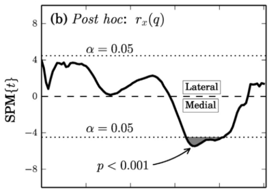
Pataky et al. (2014) JoB

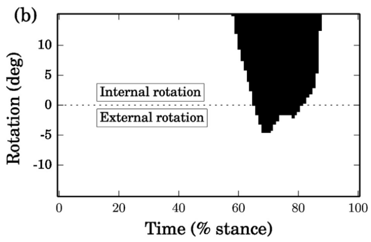
Hotelling's T² tests



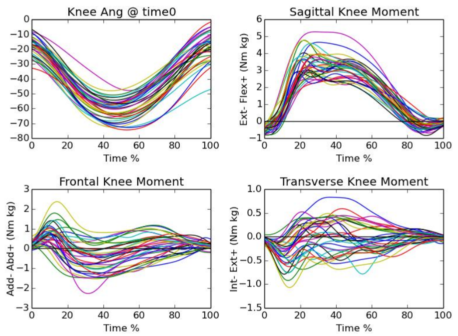
Hotelling's T² test



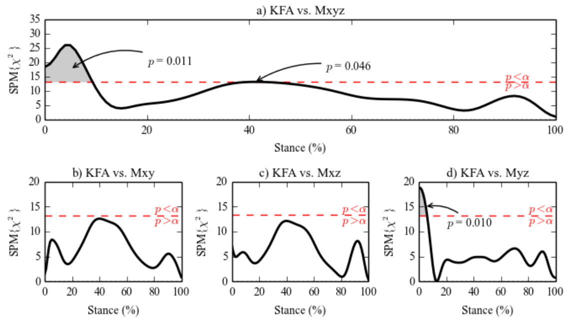




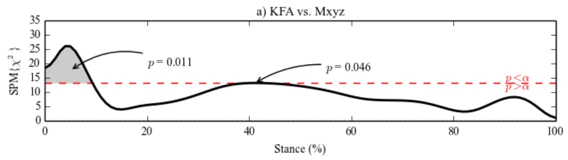
Example 2: Knee moments



CCA



CCA



Topics

- Overview
- False positives
- Vector field statistics
- Reminder of SPM limitations
- Concluding remarks

SPM limitations

- SPM has same limitations as other methods e.g. non-random sampling, non-blind experimentation, non-homologous data….
- Requires data registration (temporal normalization)
- SPM's procedures are more complex, especially RFT-based inference
- Accessibility (Python/Matlab programming)

SPM limitations

Topics

- Overview
- False positives
- Vector field statistics
- Reminder of SPM limitations
- Concluding remarks

Concluding remarks

- Support:
  - Website WWW.SPM1D.ORG
  - Handouts + notes + worksheets workshop
  - Todd, Mark, Jos
- Certificate
- Feedback