Differentiation tests for the mean shape and its variance

Stefan GIEBEL (University of Luxembourg)
joint work with Jang SCHILTZ (University of Luxembourg)
& Jens-Peter SCHENK (University of Heidelberg)

COMPSTAT 2010, Paris: August 22, 2010

Outline

1. Statistical shape analysis

Outline

1. Statistical shape analysis
2. Renal tumors in early childhood

Outline

1. Statistical shape analysis
2. Renal tumors in early childhood
3. Experimental results

Outline

Outline

Outline

Landmarks

Suppose that we want to study n objects by means of statistical shape analysis.

Landmarks

Suppose that we want to study n objects by means of statistical shape analysis.

A landmark is a point of correspondence on each object that matches between and within populations.

Landmarks

Suppose that we want to study n objects by means of statistical shape analysis.

A landmark is a point of correspondence on each object that matches between and within populations.

Denote the number of landmarks by k.

Landmarks

Suppose that we want to study n objects by means of statistical shape analysis.

A landmark is a point of correspondence on each object that matches between and within populations.

Denote the number of landmarks by k.

Every object o_i in a space V of dimension m is thus represented in a space of dimension $k \cdot m$ by a set of landmarks:

$$\forall i = 1 \dots n,\ o_i = \{l_1 \dots l_k\}, l_j \in \mathbb{R}^m. \tag{1}$$

Removing the scale

Removing the scale

1. For every $i, i = 1, ..., n$, the size of each object is determined as the euclidian norm of their landmarks.

$$\|o_i\| = \sqrt{\sum_{j=1}^{k} \|l_j^i\|_m^2}. \tag{2}$$

Removing the scale

1. For every $i, i = 1, ..., n$, the size of each object is determined as the euclidian norm of their landmarks.

$$\|o_i\| = \sqrt{\sum_{j=1}^{k} \|l_j^i\|_m^2}. \tag{2}$$

2. The landmarks are standardized by dividing them by the size of their object:

$$\tilde{l}_j^i = \frac{l_j^i}{\|o_i\|}. \tag{3}$$

Removing the location

To remove the location of the object, the landmarks are centered by the following procedure:

Removing the location

To remove the location of the object, the landmarks are centered by the following procedure:

1. For every $i, i = 1, ..., n$, we compute the the arithmetic mean z^i of the k standardized landmarks of the ith object :

$$z^i = \frac{1}{k} \sum_{j=1}^{k} \tilde{l}_j^i \tag{4}$$

Removing the location

To remove the location of the object, the landmarks are centered by the following procedure:

1. For every $i, i = 1, ..., n$, we compute the the arithmetic mean z^i of the k standardized landmarks of the ith object :

$$z^i = \frac{1}{k} \sum_{j=1}^{k} \tilde{l}_j^i \tag{4}$$

2. We center all the landmarks by subtracting this mean:

$$\bar{l}_j^i = l_j^i - z^i \tag{5}$$

Remark

We do not need to remove rotation in our application, since we use MRI images of the tumors which are frontal or transversal images.

Remark

We do not need to remove rotation in our application, since we use MRI images of the tumors which are frontal or transversal images.

We have no rotated images in our sample.

Remark

We do not need to remove rotation in our application, since we use MRI images of the tumors which are frontal or transversal images.

We have no rotated images in our sample.

Hence, we are able to work completely in the standard three-dimensional space with the euclidian norm.

Remark

We do not need to remove rotation in our application, since we use MRI images of the tumors which are frontal or transversal images.

We have no rotated images in our sample.

Hence, we are able to work completely in the standard three-dimensional space with the euclidian norm.

We do not need any further procrustes analysis nor any complicated stochastic geometry. It is easy to show that the partial procrustean distance is equivalent to the euclidean distance in our case.

The mean shape

To compare the standardized and centered sets of landmarks, we need to define the mean shape of all the objects and a distance function which allows us to evaluate how "near" every object is from this mean shape.

The mean shape

To compare the standardized and centered sets of landmarks, we need to define the mean shape of all the objects and a distance function which allows us to evaluate how "near" every object is from this mean shape.

The term "mean" is here used in the sense of Fréchet (1948).

The mean shape

To compare the standardized and centered sets of landmarks, we need to define the mean shape of all the objects and a distance function which allows us to evaluate how "near" every object is from this mean shape.

The term "mean" is here used in the sense of Fréchet (1948).

If X demotes a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ with values in a metric space (Ξ, d), an element $m \in \Xi$ is called a mean of $x_1, x_2, \ldots, x_k \in \Xi$ if

$$\sum_{j=1}^{k} d(x_j, m)^2 = \inf_{\alpha \in \Xi} \sum_{j=1}^{k} d(x_j, \alpha)^2. \qquad (6)$$

The mean shape

To compare the standardized and centered sets of landmarks, we need to define the mean shape of all the objects and a distance function which allows us to evaluate how "near" every object is from this mean shape.

The term "mean" is here used in the sense of Fréchet (1948).

If X demotes a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ with values in a metric space (Ξ, d), an element $m \in \Xi$ is called a mean of $x_1, x_2, ..., x_k \in \Xi$ if

$$\sum_{j=1}^{k} d(x_j, m)^2 = \inf_{\alpha \in \Xi} \sum_{j=1}^{k} d(x_j, \alpha)^2. \tag{6}$$

That means that the mean shape is defined as the shape with the smallest variance of all shapes in a group of objects.

The algorithm of Ziezold (1994)

To begin, we fix the mean of all the standardized and centered objects as starting value: $$\tilde{m}_0 = \frac{1}{n}\sum_{i=1}^{n} \overline{o}_i.$$

The algorithm of Ziezold (1994)

To begin, we fix the mean of all the standardized and centered objects as starting value: $$\tilde{m}_0 = \frac{1}{n}\sum_{i=1}^{n} \overline{o}_i.$$

We then undertake the following steps for $i = 1, \ldots, n$

The algorithm of Ziezold (1994)

To begin, we fix the mean of all the standardized and centered objects as starting value: $\tilde{m}_0 = \frac{1}{n}\sum_{i=1}^{n} \overline{o}_i.$

We then undertake the following steps for $i = 1, \ldots, n$

1.

$$\tilde{m} \mapsto w_i(\tilde{m}) = \begin{cases} \frac{\langle \tilde{m}, o_i \rangle}{|\langle \tilde{m}, o_i \rangle|} & \text{if } \langle \tilde{m}, o_i \rangle \neq 0 \\ 1 & \text{if } \langle \tilde{m}, o_i \rangle = 0 \end{cases} \quad (7)$$

The algorithm of Ziezold (1994)

To begin, we fix the mean of all the standardized and centered objects as starting value: $\tilde{m}_0 = \frac{1}{n}\sum_{i=1}^{n} \overline{o}_i.$

We then undertake the following steps for $i = 1, \dots, n$

1. $$\tilde{m} \mapsto w_i(\tilde{m}) = \begin{cases} \frac{\langle \tilde{m}, o_i \rangle}{|\langle \tilde{m}, o_i \rangle|} & \text{if } \langle \tilde{m}, o_i \rangle \neq 0 \\ 1 & \text{if } \langle \tilde{m}, o_i \rangle = 0 \end{cases} \quad (7)$$

2. $$\tilde{m} \mapsto T(\tilde{m}) = \frac{1}{n}\sum_{i=1}^{n} w_i(\tilde{m}) o_i \quad (8)$$

The algorithm of Ziezold (1994)

To begin, we fix the mean of all the standardized and centered objects as starting value: $\tilde{m}_0 = \frac{1}{n}\sum_{i=1}^{n} \overline{o}_i.$

We then undertake the following steps for $i = 1, \ldots, n$

1. $$\tilde{m} \mapsto w_i(\tilde{m}) = \begin{cases} \frac{\langle \tilde{m}, o_i \rangle}{|\langle \tilde{m}, o_i \rangle|} & \text{if } \langle \tilde{m}, o_i \rangle \neq 0 \\ 1 & \text{if } \langle \tilde{m}, o_i \rangle = 0 \end{cases} \quad (7)$$

2. $$\tilde{m} \mapsto T(\tilde{m}) = \frac{1}{n}\sum_{i=1}^{n} w_i(\tilde{m}) o_i \quad (8)$$

3. $$\tilde{m}_r = T(\tilde{m}_{r-1}), r = 1, 2, \ldots \quad (9)$$

The algorithm of Ziezold (1994)

To begin, we fix the mean of all the standardized and centered objects as starting value: $\tilde{m}_0 = \frac{1}{n}\sum_{i=1}^{n} \overline{o}_i.$

We then undertake the following steps for $i = 1, \ldots, n$

1. $$\tilde{m} \mapsto w_i(\tilde{m}) = \begin{cases} \frac{\langle \tilde{m}, o_i \rangle}{|\langle \tilde{m}, o_i \rangle|} & \text{if } \langle \tilde{m}, o_i \rangle \neq 0 \\ 1 & \text{if } \langle \tilde{m}, o_i \rangle = 0 \end{cases} \tag{7}$$

2. $$\tilde{m} \mapsto T(\tilde{m}) = \frac{1}{n}\sum_{i=1}^{n} w_i(\tilde{m}) o_i \tag{8}$$

3. $$\tilde{m}_r = T(\tilde{m}_{r-1}), r = 1, 2, \ldots \tag{9}$$

The stopping rule is $\tilde{m} = T(\tilde{m})$.

Outline

Renal tumors in early childhood

- Wilms-tumors (nephroblastoma) growing next to the kidney.

  Genetic cause. There are four types of tissue (a, b, c, d) and three stages of development (I, II, III).

  Many renal tumors in the childhood are diagnosed as Wilms (130 per year).

Renal tumors in early childhood

- Wilms-tumors (nephroblastoma) growing next to the kidney.

  Genetic cause. There are four types of tissue (a, b, c, d) and three stages of development (I, II, III).

  Many renal tumors in the childhood are diagnosed as Wilms (130 per year).
- Renal cell carcinoma growing also next to the kidney.

  Are rare in childhood (12 per year) but frequent for adults.

Renal tumors in early childhood

- Wilms-tumors (nephroblastoma) growing next to the kidney.

  Genetic cause. There are four types of tissue (a, b, c, d) and three stages of development (I, II, III).

  Many renal tumors in the childhood are diagnosed as Wilms (130 per year).
- Renal cell carcinoma growing also next to the kidney.

  Are rare in childhood (12 per year) but frequent for adults.
- Neuroblastoma growing next to nerve tissue.

  Quite frequent (80 per year).

Renal tumors in early childhood

- Wilms-tumors (nephroblastoma) growing next to the kidney.

  Genetic cause. There are four types of tissue (a, b, c, d) and three stages of development (I, II, III).

  Many renal tumors in the childhood are diagnosed as Wilms (130 per year).
- Renal cell carcinoma growing also next to the kidney.

  Are rare in childhood (12 per year) but frequent for adults.
- Neuroblastoma growing next to nerve tissue.

  Quite frequent (80 per year).
- Clear cell carcinoma growing next to bones.

  Rare (12 per year).

Outline

The data

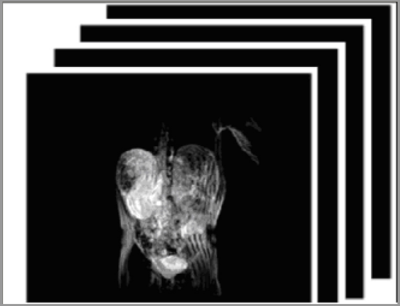
Research sample:

- Magnetic resonance images of 83 cases of tumors in frontal perspective (69 Wilms, 6 neuroblastoma, 5 clear cell carcinoma and 3 renal cell carcinoma).

The data

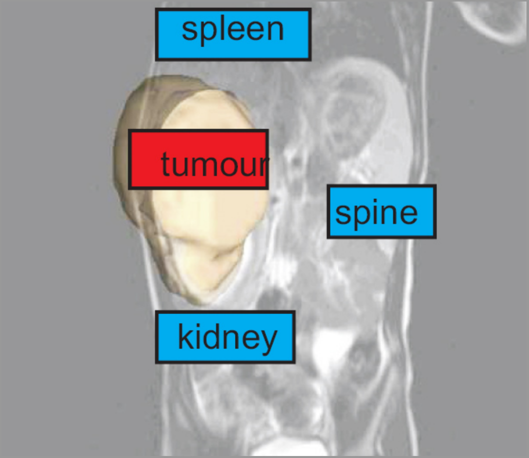
Research sample:

- Magnetic resonance images of 83 cases of tumors in frontal perspective (69 Wilms, 6 neuroblastoma, 5 clear cell carcinoma and 3 renal cell carcinoma).



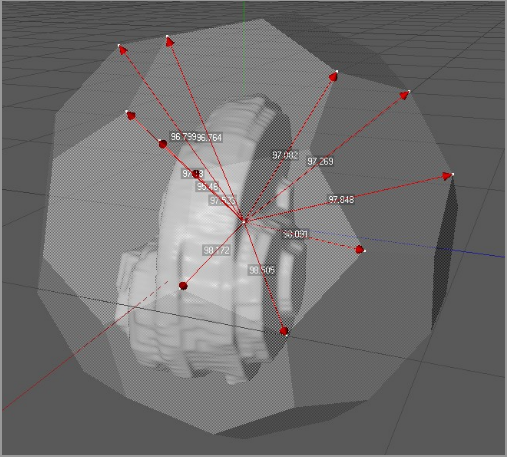
MRI image of a renal tumor in frontal view.

The three-dimensional object



Three-dimensional model of a tumor.

The platonic body C60



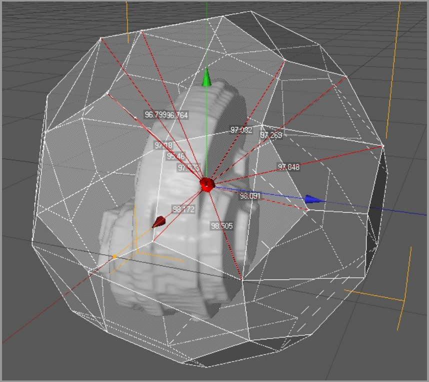
For every object, we consider the platonic body C60 whose center lies in the center of the object. This platonic body has 60 edges which give us 60 three-dimensional landmarks for every object.

The landmarks

We take as landmarks the 60 points on the border of each object closest to the edges of the platonic body.

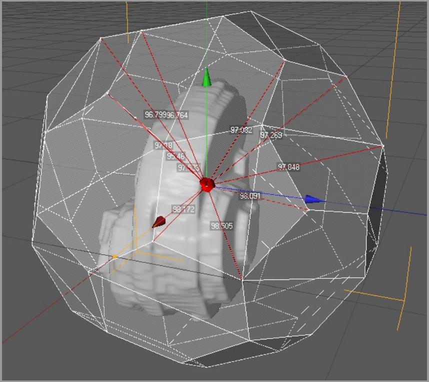
The landmarks

We take as landmarks the 60 points on the border of each object closest to the edges of the platonic body.



The landmarks

We take as landmarks the 60 points on the border of each object closest to the edges of the platonic body.



Only real measured points on the border of the tumor are taken, the approximated part of the three-dimensional object is not used.

Outline

Ziezold's test for differentiation of the types of tumors

We consider to subsets A and B of the sample of size n and $N - n$ respectively.

Ziezold's test for differentiation of the types of tumors

We consider to subsets A and B of the sample of size n and $N - n$ respectively.

The subset A is a realization of a distribution P and the subset B is an independent realization of a distribution Q.

Ziezold's test for differentiation of the types of tumors

We consider to subsets A and B of the sample of size n and $N - n$ respectively.

The subset A is a realization of a distribution P and the subset B is an independent realization of a distribution Q.

The test hypotheses are:

$$\begin{array}{ll} \textit{Hypothesis:} & H_0 : P = Q \\ \textit{Alternative:} & H_1 : P \neq Q \end{array}$$

Ziezold's test for differentiation of the types of tumors

1. Computing the mean shape m_0 of subset A.

Ziezold's test for differentiation of the types of tumors

1. Computing the mean shape m_0 of subset A.
2. Computing the u-value

Ziezold's test for differentiation of the types of tumors

1. Computing the mean shape m_0 of subset A.
2. Computing the u-value

$$u_0 = \sum_{j=1}^{n} card\big(b_k : d(b_k, m_0) < d(a_j, m_0)\big).$$

Ziezold's test for differentiation of the types of tumors

1. Computing the mean shape m_0 of subset A.
2. Computing the u-value

$$u_0 = \sum_{j=1}^{n} card\left(b_k : d(b_k, m_0) < d(a_j, m_0)\right).$$

3. Determination of all the possibilities of dividing the set into two subset with the same proportion.

Ziezold's test for differentiation of the types of tumors

1. Computing the mean shape m_0 of subset A.
2. Computing the u-value

$$u_0 = \sum_{j=1}^{n} card\left(b_k : d(b_k, m_0) < d(a_j, m_0)\right).$$

3. Determination of all the possibilities of dividing the set into two subset with the same proportion.
4. Comparing the u_0-value to all possible u-values. Computing the rank (small u-value mean a small rank).

Ziezold's test for differentiation of the types of tumors

1. Computing the mean shape m_0 of subset A.
2. Computing the u-value

$$u_0 = \sum_{j=1}^{n} card\left(b_k : d(b_k, m_0) < d(a_j, m_0)\right).$$

3. Determination of all the possibilities of dividing the set into two subset with the same proportion.
4. Comparing the u_0-value to all possible u-values. Computing the rank (small u-value mean a small rank).
5. Calculate the p-value for H_0. $p_{r=i} = \frac{1}{\binom{N}{n}}$ for $i = 1, \ldots, \binom{N}{n}$, where r is the rank for which we assume a uniform distribution.

Wilms tumors against non Wilms tumors

- **Comparing the Wilms tumors to the mean shape of the non Wilms tumors.**

Wilms tumors against non Wilms tumors

- **Comparing the Wilms tumors to the mean shape of the non Wilms tumors.**

  $u = 185 \qquad \text{rank} = 970$

  Random sample: $n = 1000 \qquad p = 0,97.$

Wilms tumors against non Wilms tumors

- **Comparing the Wilms tumors to the mean shape of the non Wilms tumors.**

  $u = 185 \qquad \text{rank} = 970$

  Random sample: $n = 1000 \qquad p = 0,97.$

- **Comparing the non Wilms tumors to the mean shape of the Wilms tumors.**

Wilms tumors against non Wilms tumors

- **Comparing the Wilms tumors to the mean shape of the non Wilms tumors.**

  $u = 185 \qquad \text{rank} = 970$

  Random sample: $n = 1000 \qquad p = 0,97.$

- **Comparing the non Wilms tumors to the mean shape of the Wilms tumors.**

  $u = 257 \qquad \text{rank} = 1 - 2$

  Random sample: $n = 1000 \qquad p = 0,002.$

Wilms tumors against Neuroblastoma

- **Comparing the Wilms tumors to the mean shape of the Neuroblastoma.**

Wilms tumors against Neuroblastoma

- **Comparing the Wilms tumors to the mean shape of the Neuroblastoma.**

  $u = 2 \qquad \text{rank} = 78$

  Random sample: $n = 1000 \qquad p = 0.078.$

Wilms tumors against Neuroblastoma

- **Comparing the Wilms tumors to the mean shape of the Neuroblastoma.**

  $u = 2 \qquad \text{rank} = 78$

  Random sample: $n = 1000 \qquad p = 0.078.$

- **Comparing the Neuroblastoma to the mean shape of the Wilms tumors.**

Wilms tumors against Neuroblastoma

- **Comparing the Wilms tumors to the mean shape of the Neuroblastoma.**

  $u = 2 \qquad \text{rank} = 78$

  Random sample: $n = 1000 \qquad p = 0.078.$

- **Comparing the Neuroblastoma to the mean shape of the Wilms tumors.**

  $u = 257 \qquad \text{rank} = 15 - 40$

  Random sample: $n = 1000 \qquad p = 0.040.$

Mean variance of a set of shapes

We define the mean variance in the sense of Fréchet of a set of objects as the average of the distances to the mean shape.
If X denotes a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ with values in a metric space (Ξ, d) and $m \in \Xi$ is the mean of $x_1, x_2, ..., x_k \in \Xi$, σ^2 is the variance of $x_1, x_2, ..., x_k \in \Xi$ if

$$\sum_{j=1}^{k} d(d(x_j, m)^2, \sigma^2)^2 = \inf_{\alpha \in \Xi} \sum_{j=1}^{k} d(d(x_j, m)^2, \alpha)^2. \qquad (10)$$

That means that the variance is defined as the mean of the distances between the "mean shape" and the objects.

The variance test

In this section we propose a test to compare the mean variance of two groups of objects.

step 1: Definition of the set of objects

There is a set M that can be divided into a subset A, realisation of a distribution P with variance σ_1^2 and a subset B, realisation of a distribution Q with variance σ_2^2.

Hypothesis: $H_0 : \sigma_1^2 = \sigma_2^2$
Alternative: $H_1 : \sigma_1^2 \neq \sigma_2^2$

The variance test

step 2: Computation of the variance

The variance is calculated by means of a straightforward generalisation of the algorithm of Ziezold (1994).

step 3: Computation of the F-value

$$F = \frac{|\hat{\sigma}_1^2|}{|\hat{\sigma}_2^2|}.$$

step 4: Determination of all the possibilities of dividing the set into two subsets with given sizes

step 5: Comparison of the F-value to all possible F-values. Computation of the rank (small F-value mean a small rank).

The variance test

step 6: Computation of the p-value for H_0

$p_{r=i} = 1 - \frac{1}{\binom{N}{n}}$ for $i = 1, \ldots, \binom{N}{n}$, where r is the rank for which we assume a rectangular distribution on the right side and $p_{r=i} = \frac{1}{\binom{N}{n}}$ on the left side.

Results of variance test

For the renal tumours, the F-value for the differentiation of the variance of the group of nephroblastomas to the group of neuroblastomas is 1.28128 and the rank is 315.
So the corresponding p-value is $1 - 0.315 = 0.685$ and we have to accept the null hyptothesis that the variance is similiar in both groups.
Both kind of tumours seem to have more or less the same dispersion and a possible difference in the dispersion can be excluded as cause for difficulties in distinguishing the two kinds of tumours.

Outline

Conclusion

Three-dimensional statistical shape analysis seems to be a good tool for differentiating the renal tumors appearing in early childhood.

Conclusion

Three-dimensional statistical shape analysis seems to be a good tool for differentiating the renal tumors appearing in early childhood.

- Wilms tumors can be clearly differentiated from neuroblastomas.

Conclusion

Three-dimensional statistical shape analysis seems to be a good tool for differentiating the renal tumors appearing in early childhood.

- Wilms tumors can be clearly differentiated from neuroblastomas.
- It is possible to differentiate the whole set of non-Wilms tumors from the mean shape of Wilms tumors.

Conclusion

Three-dimensional statistical shape analysis seems to be a good tool for differentiating the renal tumors appearing in early childhood.

- Wilms tumors can be clearly differentiated from neuroblastomas.
- It is possible to differentiate the whole set of non-Wilms tumors from the mean shape of Wilms tumors.
- But we cannot use statistical shape analysis to say if a given general tumor is not a Wilms tumor.

Conclusion

Three-dimensional statistical shape analysis seems to be a good tool for differentiating the renal tumors appearing in early childhood.

- Wilms tumors can be clearly differentiated from neuroblastomas.
- It is possible to differentiate the whole set of non-Wilms tumors from the mean shape of Wilms tumors.
- But we cannot use statistical shape analysis to say if a given general tumor is not a Wilms tumor.
- The variance of the set of Wilms tumors is not significantly different from the variance of the set of the other tumors.