About Database

The database contains information about complete population of the Silesian Horses bred in Poland after 1945 year and consists of 18,980 objects (13,408 objects describing mares, and 5572 describing stallions). Each horse is depict by 41 attributes which can be divided into three groups:

- pedigree features – such as horse id, date of birth, breeder, father and mother id, inbreed, groups of ancestors, etc.;
- zoometric features – such as height at whiters, chest circumference, massiveness index, etc.;
- point–scale estimation features – such as bonitation points for build, type, foundation, etc.

Bellow u can find detailed description of each attribute.

Attributes

Pedigree features:

- **Horse Id** – Id of the horse:
  - 18,980 distinct values;
  - 0% missing data;
- **Horse Id2** – Id of the horse (alternative numeration):
  - 18,980 distinct values;
  - 0% missing data;
- **Sex** – sex of the horse:
  - **O** – stands for stallions;
  - **K** – stands for mares;
  - 2 distinct values;
  - 0% missing data;
- **Date of Birth** – date of birth of the horse:
  - 7353 distinct values;
  - 0% missing data;
- **Year of Birth** – year of birth of the horse:
  - 67 distinct values;
  - 0% missing data;
- **Breeder** – breeder of the horse (geographical location):
  - **1pan** – state horse-breeding farms;
  - **2wro** – Wroclaw region;
  - **3kat** – Katowice and Opole region;
  - **4kie** – Kielce region;
  - **5rze** – Rzeszow region;
  - **6nie** – non-Silesian regions (rest of the Poland);
  - **7obc** – foreign breeders (imported horses, ancestors of foreign races);
  - 7 distinct values;
  - 0% missing data;

- **Decade** – decades of horse breeding:
  - **4** – years to 1949;
  - **5** – years from 1950 to 1959;
  - **6** – years from 1960 to 1969;
  - **7** – years from 1970 to 1979;
  - **8** – years from 1980 to 1989;
  - **9** – years from 1990 to 1999;
  - **10** – years from 2000 till now;
  - 7 distinct values;
  - 0% missing data;
- **Breed Decade** – decades of horse breeding program:
  - **D0** – years to 1946;
  - **D1** – years from 1947 to 1958;
  - **D2** – years from 1959 to 1968;
  - **D3** – years from 1969 to 1982;
  - **D4** – years from 1983 to 1996;
  - **D5** – years from 1997 till now;
  - 6 distinct values;
  - 0% missing data;
- **Inbreed** – different groups(factors) of inbreeding:
  - **1** – unrelated;
  - **2** – from 0,1 to 1 group;
  - **3** – from 1,1 to 4 group;
  - **4** – above 4 group;
  - 4 distinct values;
  - 0% missing data;
- **Group of others ancestors** – blood involvement of other races:
  - **1** – no involvement;
  - **2** – to 12,5% of involvement;
  - **3** – from 12,6% to 49% involvement;
  - **4** – above 50% of involvement;
  - 4 distinct values;
  - 0% missing data;
  - Minimal value: 1;
  - Maximal value: 4;
  - Mean value: 1,679;
- **Percent of others ancestors** – percentage of blood involvement of other races:
  - 4 distinct values;
  - 63% missing data;
  - Minimal value: 6,7;
  - Maximal value: 100;
  - Mean value: 27,996;
- **Group of Thoroughbred ancestors** – blood involvement of Thoroughbred races:
  - **1** – no involvement;
  - **2** – to 12,5% of involvement;
  - **3** – from 12,6% to 49% involvement;

    - **4** – above 50% of involvement;
    - 4 distinct values;
    - 0% missing data;
    - Minimal value: 1;
    - Maximal value: 4;
    - Mean value: 1,341;
- **Percent of Thoroughbred ancestors** – percentage of blood involvement of Thoroughbred races:
    - 4 distinct values;
    - 82% missing data;
    - Minimal value: 12,5
    - Maximal value: 87,5
    - Mean value: 28,803
- **Group of Schweres Warmblut ancestors** – blood involvement of Schweres Warmblut races:
    - **1** – no involvement;
    - **2** – to 12,5% of involvement;
    - **3** – from 12,6% to 49% involvement;
    - **4** – above 50% of involvement;
    - 4 distinct values;
    - 0% missing data;
    - Minimal value: 1;
    - Maximal value: 4;
    - Mean value: 1,373;
- **Percent of Schweres Warmblut ancestors** – percentage of blood involvement of Schweres Warmblut races:
    - 4 distinct values;
    - 76% missing data;
    - Minimal value: 6,7
    - Maximal value: 87,5
    - Mean value: 21,639
- **Sire Lines/Breed Groups –** ancestor lines or breed groups of the horse:
    - Sire Lines:
        - **Auto** – line ancestor: Automat;
        - **Brun** – line ancestor: Bruno;
        - **Dieb** – line ancestor: Diebitsch;
        - **Fab** – line ancestor: Fabian;
        - **Firl –** line ancestor: Firley;
        - **Hold –** line ancestor: Holdek;
        - **Prud –** line ancestor: Prudnik;
        - **Ulan –** line ancestor: Ulan;
    - Breed Groups:
        - **alt** – pre-war German horses;
        - **slas** – rest of the Silesian horses (horses without documented origin);
        - **sp** – descendents from the half-blood ancestors;
        - **sw** – descendents from the Schweres Warmblut ancestors;

        - **xx** – descendents from the Thoroughbred ancestors;
    - 4 distinct values;
    - 76% missing data;
- **Mother Id –** dam Id:
    - 5435 distinct values;
    - 39% missing data;
- **Mother Id2** – dam Id (alternative numeration):
    - 5434 distinct values;
    - 39% missing data;
- **Fahter Id** – sire Id:
    - 2161 distinct values;
    - 7% missing data;
- **Father Id2** – sire Id (alternative numeration):
    - 2159 distinct values;
    - 7% missing data.

Zoometric features:

- **Height at Whiters** – height of the horse measured at whiters:
    - 35 distinct values;
    - 42% missing data;
    - Minimal value: 140
    - Maximal value: 176
    - Mean value: 158,634
- **Chest Circumference**:
    - 67 distinct values;
    - 42% missing data;
    - Minimal value: 157
    - Maximal value: 240
    - Mean value: 195,459
- **Fore Cannon Circumference:**
    - 32 distinct values;
    - 42% missing data;
    - Minimal value: 18;
    - Maximal value: 27;
    - Mean value: 22,234;
- **Massiveness Index –** quotient of chest circumference and height of the horse:
    - $massiveness\ index = \frac{chest\ circumference}{height\ at\ whiters}$;
    - 299 distinct values;
    - 42% missing data;
    - Minimal value: 101,9;
    - Maximal value: 145,4;
    - Mean value: 123,203;
- **Boniness Index –** quotient of fore cannon circumference and height of the horse:
    - $boniness\ index = \frac{fore\ cannon\ circumference}{height\ at\ whiters}$;

- 266 distinct values;
- 42% missing data;
- Minimal value: 11,59;
- Maximal value: 16,98;
- Mean value: 14,017;

- **Baron's Index –** quotient of chest circumference raised to the second power and height of the horse:
  - $massiveness\ index = \frac{(chest\ circumference\ )^2}{height\ at\ whiters}$;
  - 616 distinct values;
  - 42% missing data;
  - Minimal value: 161,1;
  - Maximal value: 347;
  - Mean value: 241,126;
  - In *.arff database denoted as "**Barons Index**".

Point-scale estimation features:

- **Sum of point-scale estimation** – sum of points for all bonitation features:
  - 33 distinct values;
  - 44% missing data;
  - Minimal value: 58;
  - Maximal value: 90;
  - Mean value: 75,082;
- **Build** – sum of points for "**Head and Neck**", "T**runk**", "**Type**" attributes and an overall appearance:
  - 15 distinct values;
  - 50% missing data;
  - Minimal value: 21;
  - Maximal value: 35;
  - Mean value: 29,561;
- **Sum for Limbs and Hooves** – sum of points for limbs and hooves:
  - 13 distinct values;
  - 50% missing data;
  - Minimal value: 14;
  - Maximal value: 26;
  - Mean value: 19,048;
- **Foundation** – sum of points for "**Sum for Limbs and Hooves**" attribute and movement:
  - 15 distinct values;
  - 50% missing data;
  - Minimal value: 26;
  - Maximal value: 40;
  - Mean value: 32,442;
- **Type –** points for conformity with breed:
  - 15 distinct values;
  - 50% missing data;
  - Minimal value: 8;

    - Maximal value: 15;
    - Mean value: 12,949;
- **Head and Neck** – sum of points for head and neck:
    - 4 distinct values;
    - 50% missing data;
    - Minimal value: 2;
    - Maximal value: 5;
    - Mean value: 3,666;
- **Trunk** – sum of points for trunk:
    - 8 distinct values;
    - 50% missing data;
    - Minimal value: 8;
    - Maximal value: 15;
    - Mean value: 12,946;
- **Forelegs** – sum of points for forelegs:
    - 6 distinct values;
    - 50% missing data;
    - Minimal value: 4;
    - Maximal value: 9;
    - Mean value: 6,207;
- **Hind legs** – sum of points for hind legs:
    - 5 distinct values;
    - 50% missing data;
    - Minimal value: 4;
    - Maximal value: 8;
    - Mean value: 6,232;
- **Hooves** – sum of points for hooves:
    - 6 distinct values;
    - 50% missing data;
    - Minimal value: 4;
    - Maximal value: 9;
    - Mean value: 6,609;
- **Movement** - sum of points for movement:
    - 8 distinct values;
    - 50% missing data;
    - Minimal value: 10;
    - Maximal value: 17;
    - Mean value: 13,394;
- **Appearance** – sum of points for appearance:
    - 37 distinct values;
    - 44% missing data;
    - Minimal value: 8;
    - Maximal value: 90;
    - Mean value: 20,24.