SKB content details

June 17, 2016

This document summarises the contents of the SKB, by objective. The objective is directly mapped to the type of the dependent variable within the research question or hypothesis. Three objectives are currently covered in this document (time to event, interval and nominal).

1 Time to event - S

Models:

1. KM (Kaplan Meier) $m_{s1} = km$
2. PH (Cox Proportional Hazards) $m_{s2} = ph$
3. Weibull $m_{s3} = w$
4. *to add*: competing risk models and survival forests)

Assumptions:

1. Non informative censoring a_1
2. Testing for proportional Hazards a_2
3. Testing for Weibull distribution a_3

Mapping of assumptions critical to models

- $m_{s1} = km \mapsto a_1$
- $m_{s2} = ph \mapsto a_1, m_{s2} = ph \mapsto a_2$
- $m_{s3} = w \mapsto a_1, m_{s3} = w \mapsto a_3$

Context domains:

- Censoring
- Model Intent

2 Interval I

Models:

1. anova$m_{i1} = anova$
2. t-test $m_{i2} = t$
3. Welch $m_{i3} = welch$
4. *to add* non parametric approach

Assumptions:

1. Is the independent variable normally distributed? a_4
2. Independent variable (or covariate of interest) is nominal? a_5
3. independent variable (or covariate of interest) is binary? a_6
4. Is the variance equal within each level of the target variable? (Homoscedasticy) a_7

Mapping of assumptions critical to models

- $m_{i1} = anova \mapsto a_4, m_{i1} = anova \mapsto a_7$
- $m_{i2} = t \mapsto a_4, m_{i2} = t \mapsto a_6, m_{i2} = t \mapsto a_7$
- $m_{i3} = welch \mapsto a_4, m_{i3} = welch \mapsto a_6$

Context Domains

- Missing data
- Model intent

t.test in R

```
t.test(analysis.data$age~analysis.data$Gender, var.equal=TRUE)
```

welch in R

```
t.test(analysis.data$age~analysis.data$Gender)
```

anova in R

```
fit<-aov(analysis.data$age ~ analysis.data$Gender, data=analysis.data)
summary(fit)
```

testing a_4 in R

```
shapiro.test(as.numeric(analysis.data$age))
```

testing a_7 in R

```
bartlett.test(analysis.data$age~analysis.data$Gender)
```

3 Categorical - N

Models:

1. chi squared $m_{b1} = \chi^2$
2. Fisher's exact $m_{b2} = Fisher$
3. Logistic regression $m_{b3} = lr$
4. Decision Tree $m_{b5} = dt$
5. Neural Network - details to follow

Assumptions:

1. Is the dependent variable binary? a_8
2. Are there more than 1000 cases? a_9
3. Are there more than 5 cases in one cell? a_{10}
4. Is there more than one covariate of interest? a_{11}

Mapping of assumptions critical to models

- $m_{b1} = \chi^2 \mapsto a_{10}, m_{b1} = \chi^2 \mapsto \neg a_{11}$
- $m_{b2} = Fisher \mapsto \neg a_{11}$
- $m_{b3} = lr \mapsto a_{11}, m_{b3} = lr \mapsto a_8$
- $m_{b5} = dt \mapsto a_9, m_{b5} = dt \mapsto a_{11}$

Context domains

- Model intent
- Missing data