UNIVERSITY OF CAMBRIDGE

The Psychometrics Centre

Summer School in Applied Psychometric Principles

Peterhouse College

13th to 17th September 2010

UNIVERSITY OF CAMBRIDGE

The Psychometrics Centre

Differential Item Functioning in R: The difR package

Day 5

Jan R. Böhnke
University of Trier, Germany

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Topics already covered

- Differential Item Functioning defined as systematic differences in item parameters between different subgroups of the population

DIF in R

- DIF analyses in the Rasch framework were already covered during the Rasch day:
  - Andersen Likelihood test to test whether the whole scale functions the same way in the sub-groups
  - Wald Test to identify whether single items have different difficulty parameters in the sub-groups

Data, dichotomous case

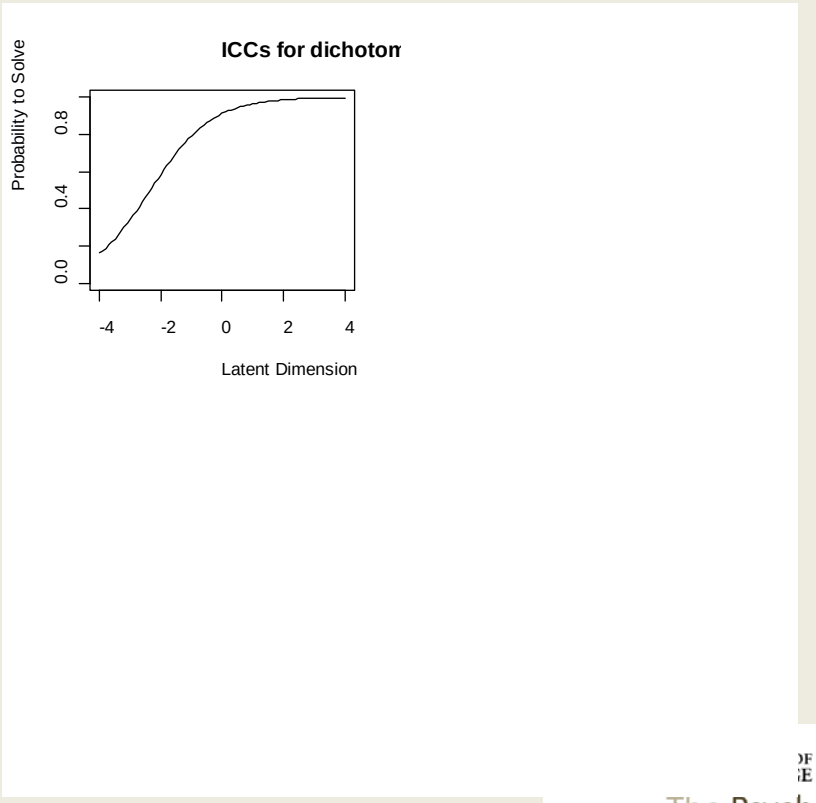
- using the simulated data from Anna‘s presentation yesterday
- results in DIFAS: item 4, 6, 10, 12, 15

DIF STATISTICS: DICHOTOMOUS ITEMS

Name	MH CHI	MH LOR	LOR SE	LOR Z	BD	CDR	ETS
Var 1	0.2461	0.0958	0.1659	0.5775	0.49	OK	A
Var 2	7.658	0.3946	0.1393	2.8327	0.365	Flag	A
Var 3	1.8162	-0.2007	0.1413	-1.4204	0.007	OK	A
Var 4	32.4658	-0.775	0.1374	-5.6405	0.122	Flag	C
Var 5	0.0342	-0.0297	0.1208	-0.2459	0.047	OK	A
Var 6	82.8232	0.9966	0.1109	8.9865	0.47	Flag	C
Var 7	0.3814	-0.0713	0.1062	-0.6714	0.484	OK	A
Var 8	0.6644	-0.0898	0.1035	-0.8676	0.393	OK	A
Var 9	4.9067	-0.2356	0.104	-2.2654	0.033	OK	A
Var 10	31.2327	0.6469	0.1151	5.6203	0.204	Flag	B
Var 11	5.8599	-0.2769	0.1119	-2.4745	2.238	Flag	A
Var 12	33.0494	-0.6519	0.1137	-5.7335	6.947	Flag	C
Var 13	1.9575	-0.1794	0.1225	-1.4645	0.583	OK	A
Var 14	5.0798	-0.2983	0.1286	-2.3196	0.093	Flag	A
Var 15	24.6969	0.7288	0.1458	4.9986	0.003	Flag	C

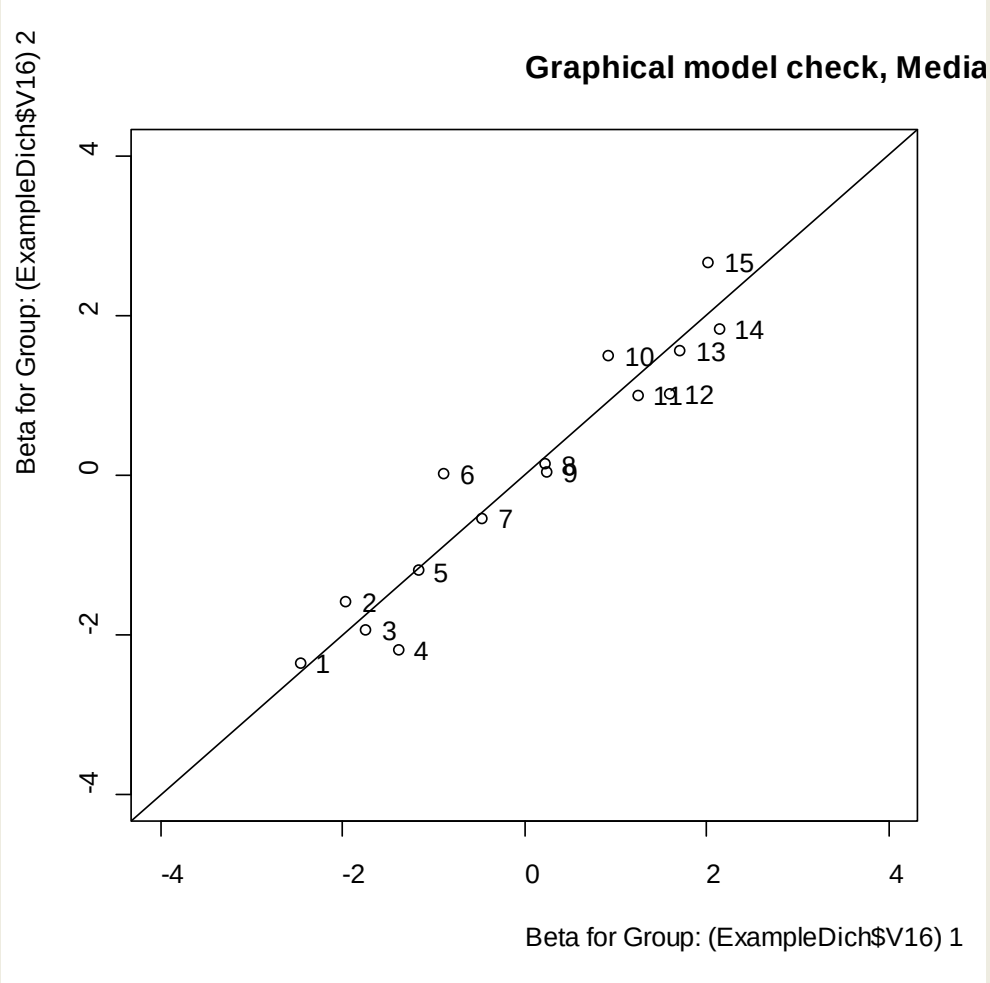
Data, dichotomous case; eRm

- using eRm; step 1: fitting the Rasch Model;
- resulting curves and confidence intervals:



Data, dichotomous case; eRm

- using eRm; step 2: Andersen Test
- LR (df=14) = 224.303; p < .001
- items: 4, 6, 10, 12, 15
- but also general more misfit
- remember: not simulated according to this model!



Data, dichotomous case; eRm

- using eRm; step 3: Wald Test
- corrected alpha level (Sidak procedure): .0034
- items 2, 4, 6, (9), 10, (11), 12, (14), 15

Wald test on item level (z-values):

	z-statistic	p-value
beta V1	0.707	0.479
beta V2	3.030	0.002
beta V3	-1.391	0.164
beta V4	-6.055	0.000
beta V5	-0.248	0.804
beta V6	8.875	0.000
beta V7	-0.709	0.478
beta V8	-0.929	0.353
beta V9	-2.123	0.034
beta V10	5.480	0.000
beta V11	-2.363	0.018
beta V12	-5.480	0.000
beta V13	-1.333	0.183
beta V14	-2.467	0.014
beta V15	4.815	0.000

DIF in R in difR

- difR is a package that provides several opportunities to calculate *dichotomous* DIF
- it is connected to the ltm package with has also be installed but no calls on that have to be made (all done by difR)

Mantel-Haenszel in difR

- command:
- mantelHaenszel()
- will need grouping vector only with 0 = reference, 1 = focal:
  - grouping<-(ExampleDich[,16]-1)
- will need object containing only the items:
  - items<-ExampleDich[,1:15]

Mantel-Haenszel in difR

- command:
- resMH1<-mantelHaenszel(items,grouping)

Mantel-Haenszel in difR

- Results from R

Data Editor

	resMH1.resMH
1	0.2460909
2	7.658029
3	1.816225
4	32.46579
5	0.03419603
6	82.82315
7	0.3814386
8	0.6644087
9	4.906731
10	31.23271
11	5.859884
12	33.04937
13	1.957509
14	5.079773
15	24.69691
16	

and from DIFAS

```
DIF STATISTICS: DICHO
---------------------
Name              MH CHI
---------------------
Var 1              0.2461
Var 2               7.658
Var 3              1.8162
Var 4             32.4658
Var 5              0.0342
Var 6             82.8232
Var 7              0.3814
Var 8              0.6644
Var 9              4.9067
Var 10            31.2327
Var 11             5.8599
Var 12            33.0494
Var 13             1.9575
Var 14             5.0798
Var 15            24.6969
---------------------
```

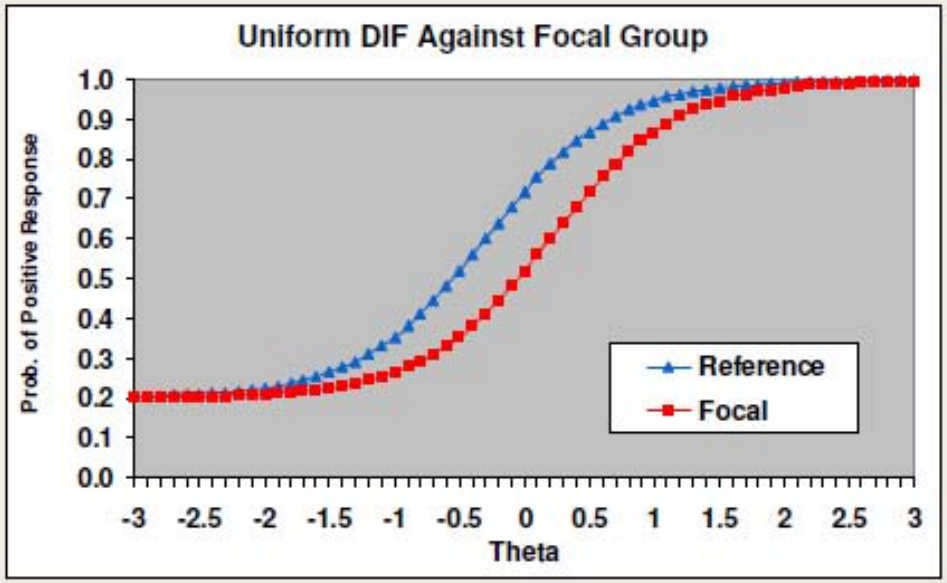
Mantel-Haenszel in difR

- the routine drops the item under consideration in that moment (as well as DIFAS as the results show)
- you can also define purification, i.e. which items should be used to represent the latent variable

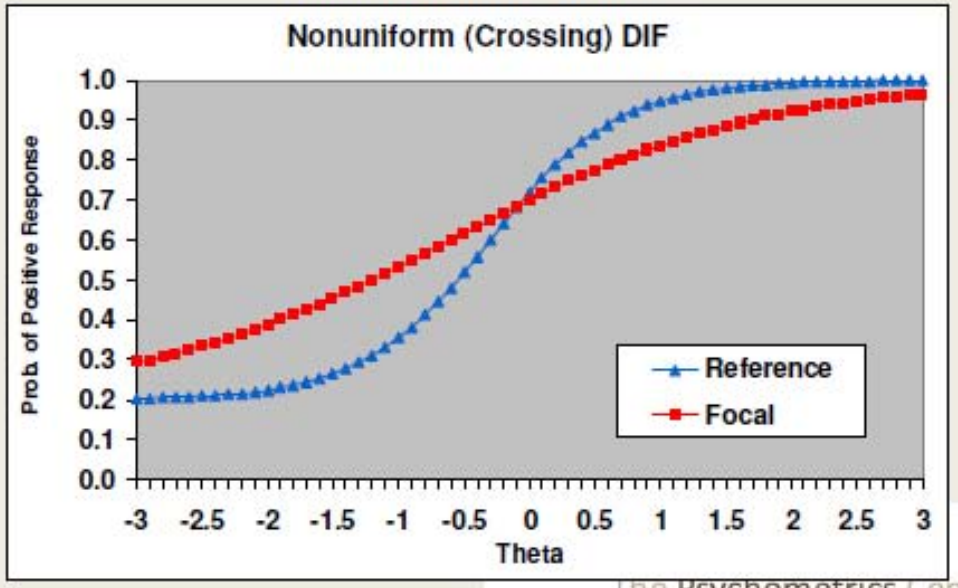
LOGISTIC REGRESSION TO DETECT DIF

UNIVERSITY OF CAMBRIDGE
The Psychometrics Centre

Uniform and non-uniform DIF

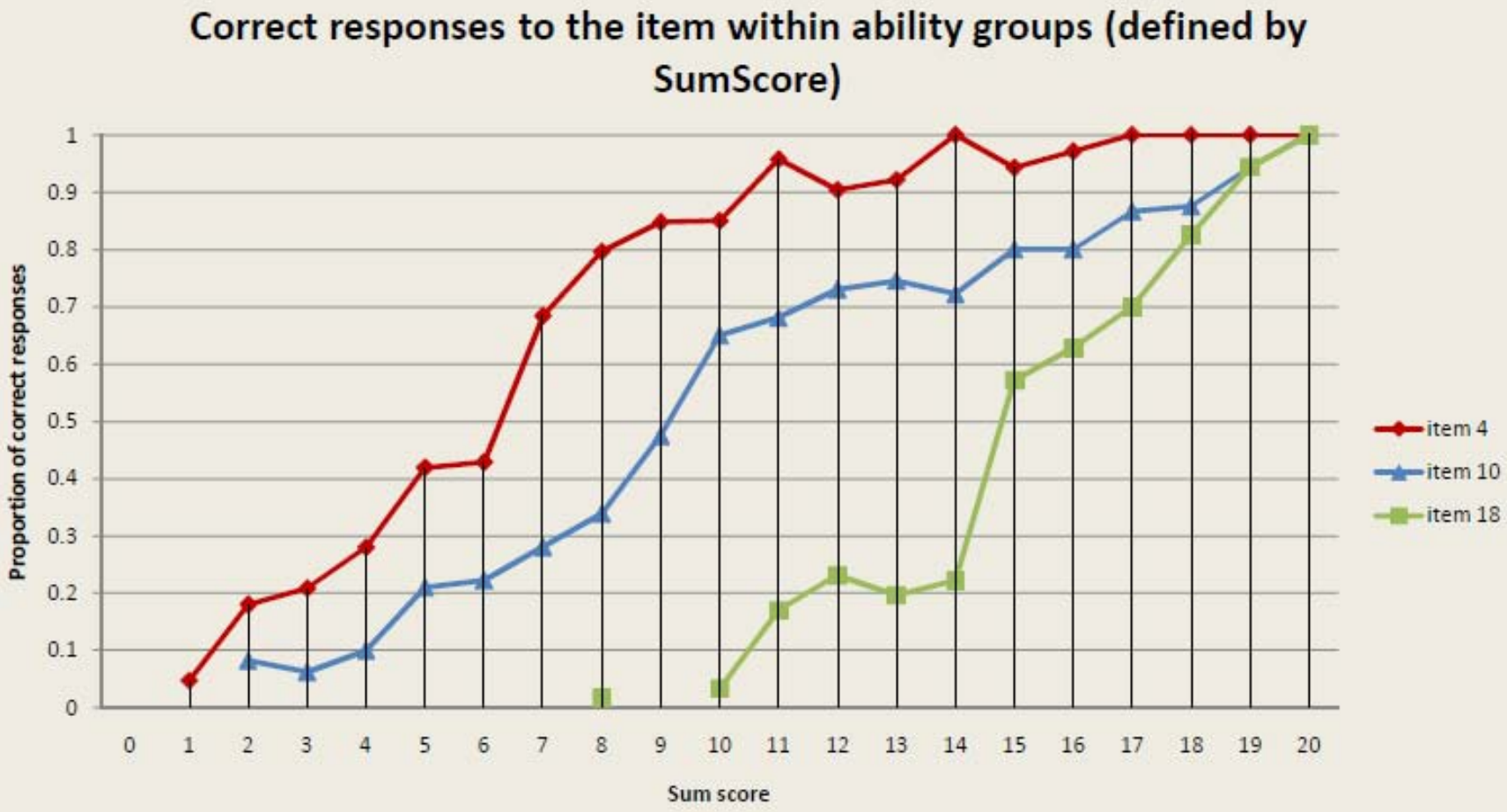


Focal group has lower probability of endorsing the item at all trait levels

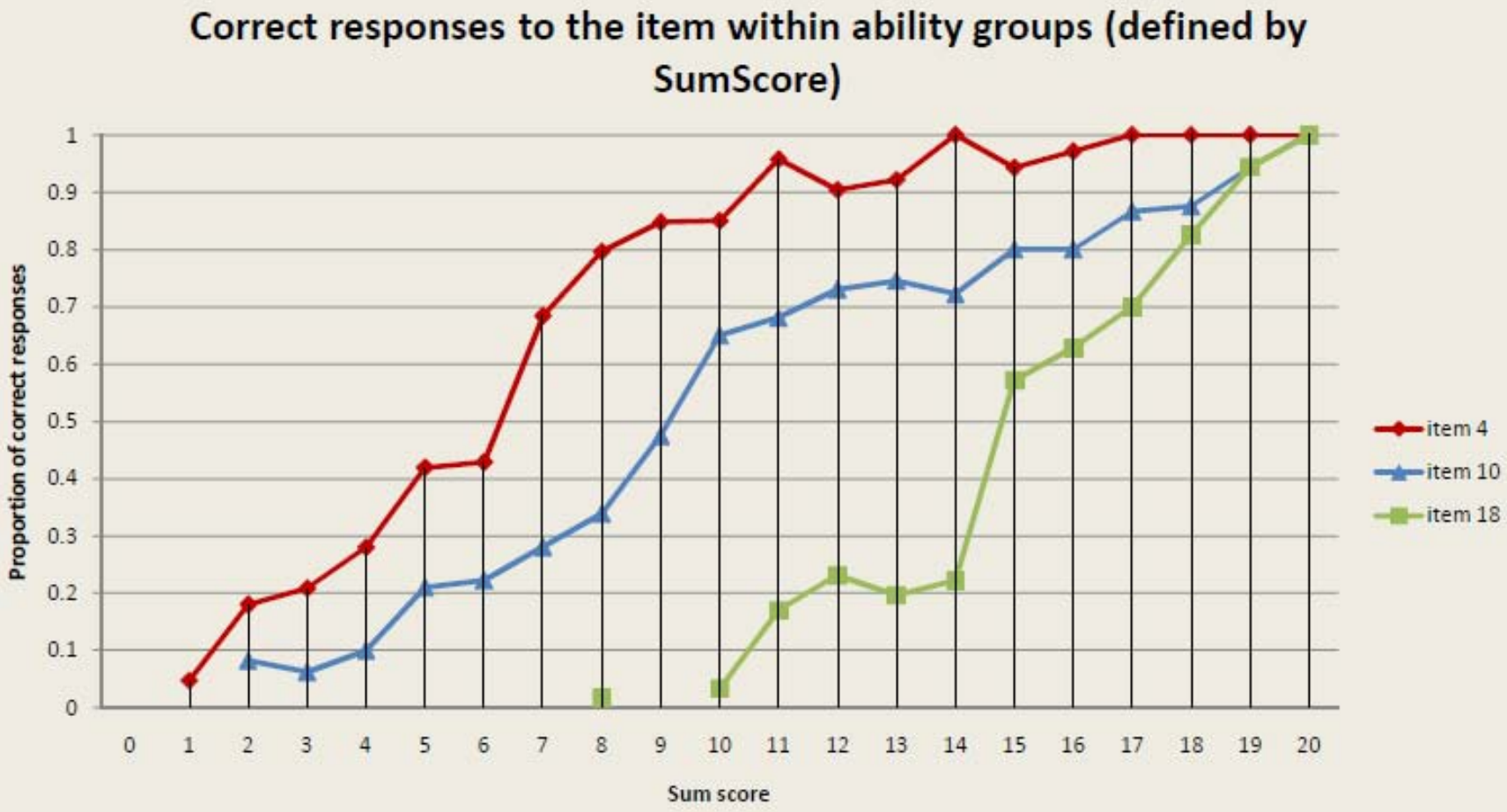


Focal group has higher probability of endorsing the item at low level of trait, but lower probability at high level

What can be said about these items?



What can be said about these items?



Logistic Regression to detect DIF

- it is assumed that you have a representation of the latent trait
  - sum score, estimate of ability from IRT model…
- empirical relative frequencies of endorsing an item depending on this proxy for the latent trait should show an approximation of the item characteristic curve

Logistic Regression to detect DIF

- these two claims say, we could run a logistic regression that predicts the probability to solve an item from the level of the latent trait

Logistic Regression to detect DIF

- this should (re-)produce our well known ICC
- if we use group in this regression, it is only significant in case of UNIFORM DIF

Logistic Regression to detect DIF

- if no uniform DIF was present, the test whether the grouping variable explains additional information beyond the score should not be significant

Logistic Regression to detect DIF

- dichUDIF<-difLogistic(items,grouping,type="udif",focal.name=1)

Logistic Regression to detect DIF

- reported is the result of the Likelihood Ratio test for every item whether adding „grouping“ as a predictor has a statistical significant effect on predicting the probability of solving the item

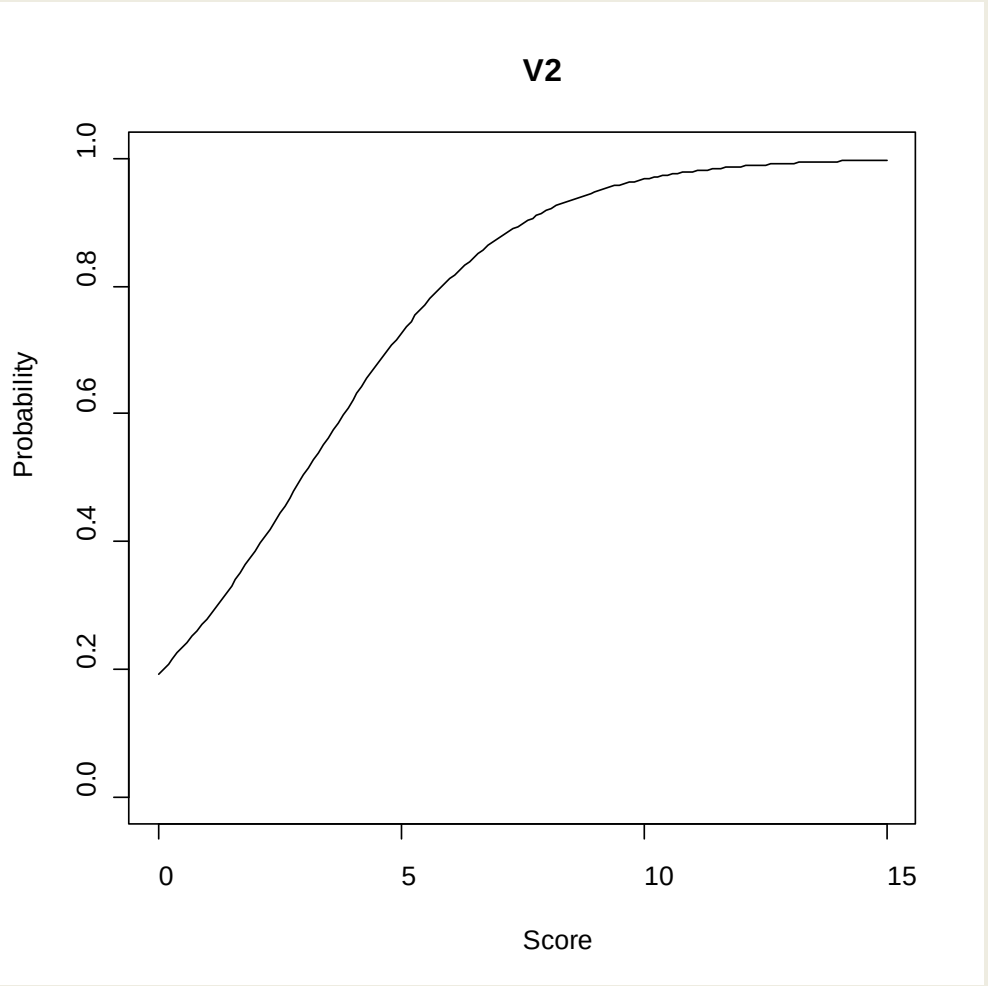
R Console

```
Detection of uniform Differential I
using Logistic regression method, w
and with LRT DIF statistic

Logistic regression DIF statistic:

     Stat.   P-value
V1    0.4684  0.4937
V2    9.1764  0.0025 **
V3    2.1723  0.1405
V4   36.5388  0.0000 ***
V5    0.0996  0.7523
V6   84.0787  0.0000 ***
V7    0.4577  0.4987
V8    0.9172  0.3382
V9    4.8994  0.0269 *
V10  31.8524  0.0000 ***
V11   6.6677  0.0098 **
V12  32.4617  0.0000 ***
V13   2.4271  0.1193
V14   6.0825  0.0137 *
V15  23.5556  0.0000 ***
```

Logistic Regression to detect DIF

- plotting of results

- plot(dichUDIF,plot="itemCurve",item=2)

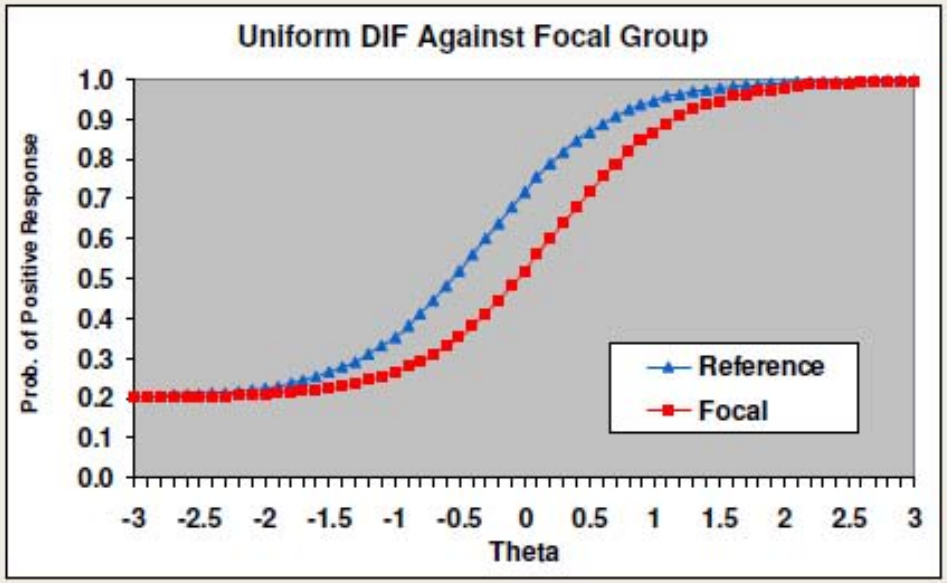


Logistic Regression to detect DIF

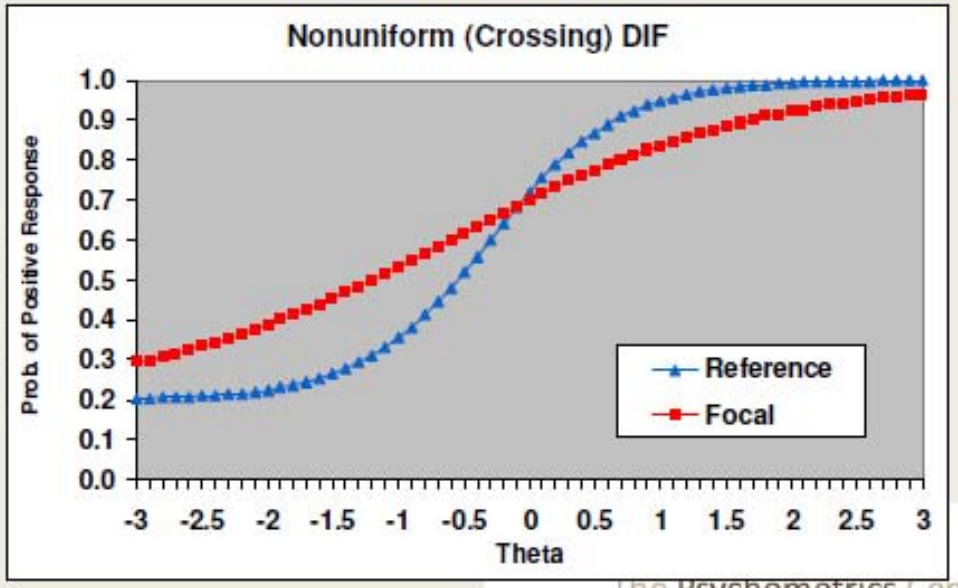
- usually identification based on effect size (Nagelkerke's Pseudo-r^2):

```
Effect size code:
 'A': negligible effect
 'B': moderate effect
 'C': large effect

    R^2    ZT JG
V1  0.0019 A  A
V2  0.0230 A  A
V3  0.0056 A  A
V4  0.1071 A  C
V5  0.0002 A  A
V6  0.1187 A  C
V7  0.0009 A  A
V8  0.0014 A  A
V9  0.0074 A  A
V10 0.0455 A  B
V11 0.0103 A  A
V12 0.0548 A  B
V13 0.0042 A  A
V14 0.0163 A  A
V15 0.0602 A  B

Effect size codes:
 Zumbo & Thomas (ZT): 0 'A' 0.13 'B' 0.26 'C' 1
 Jodoign & Gierl (JG): 0 'A' 0.035 'B' 0.07 'C' 1
```

Uniform and non-uniform DIF



Focal group has lower probability of endorsing the item at all trait levels



Focal group has higher probability of endorsing the item at low level of trait, but lower probability at high level

Logistic Regression to detect DIF

- non-uniform DIF adds to this only the interaction between grouping (G) and the trait level (T)

Logistic Regression to detect DIF

- if this interaction terms adds significant as well as relevant information compared to the uniform DIF, it is flagged

Logistic Regression to detect DIF

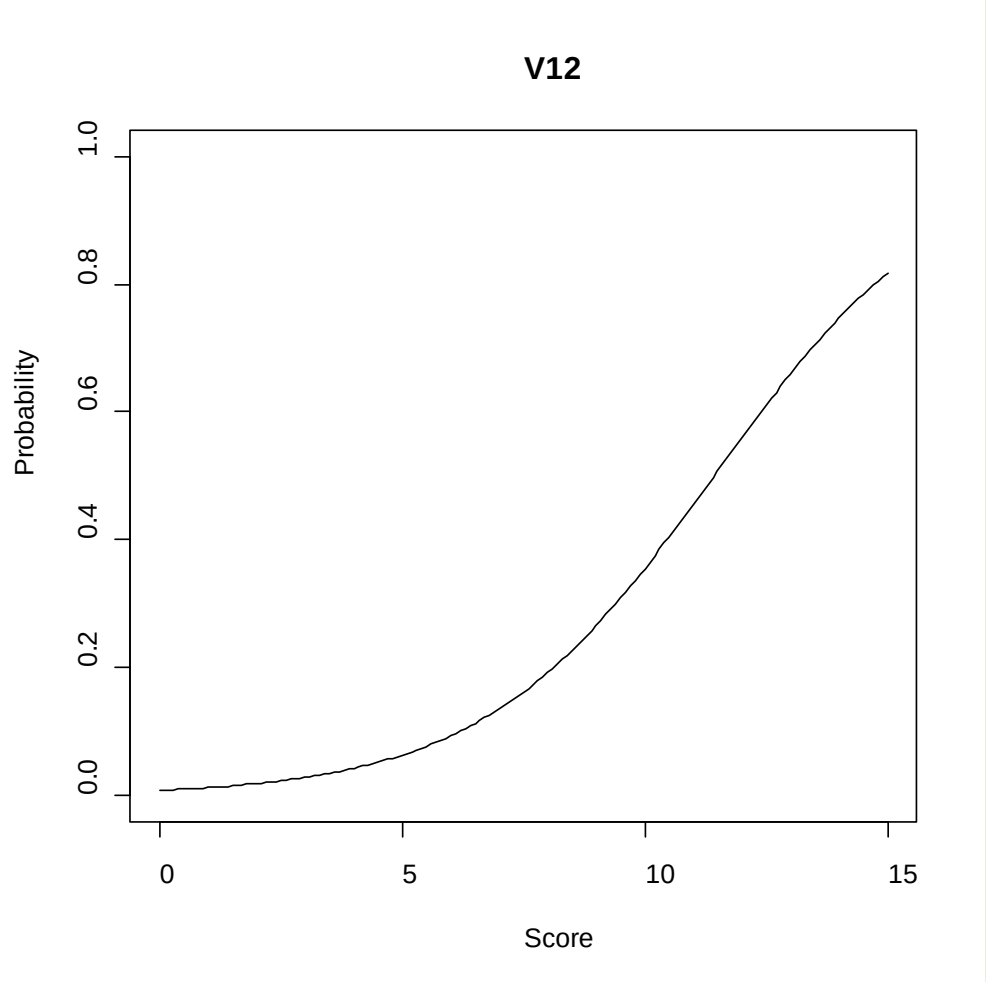
- reported is the result of the Likelihood Ratio test for every item whether adding the interaction „TG“ as a predictor has a statistical significant effect on predicting the probability of solving the item

```
Logistic regression DIF statistic:

     Stat.  P-value
V1   0.5156 0.4727
V2   0.4112 0.5214
V3   0.0216 0.8832
V4   0.3676 0.5443
V5   0.0725 0.7877
V6   0.3944 0.5300
V7   0.5056 0.4770
V8   0.1483 0.7001
V9   0.0939 0.7593
V10  0.0854 0.7701
V11  1.2723 0.2593
V12  4.9782 0.0257  *
V13  0.7529 0.3856
V14  0.0644 0.7997
V15  0.0270 0.8694
```

Logistic Regression to detect DIF

- plotting of results

- plot(dichNonUDIF, plot="itemCurve",item=12)



Logistic Regression to detect DIF

- usually identification based on effect size (Nagelkerke's Pseudo-r^2):

```
Effect size code:
 'A': negligible effect
 'B': moderate effect
 'C': large effect

    R^2    ZT JG
V1  0.0020 A  A
V2  0.0010 A  A
V3  0.0001 A  A
V4  0.0011 A  A
V5  0.0001 A  A
V6  0.0005 A  A
V7  0.0010 A  A
V8  0.0002 A  A
V9  0.0001 A  A
V10 0.0001 A  A
V11 0.0020 A  A
V12 0.0083 A  A
V13 0.0013 A  A
V14 0.0002 A  A
V15 0.0001 A  A

Effect size codes:
 Zumbo & Thomas (ZT): 0 'A' 0.13 'B' 0.26 'C' 1
 Jodoign & Gierl (JG): 0 'A' 0.035 'B' 0.07 'C' 1
```

Comparison

- Logistic regression of the probability of the item responses tests the hypotheses the IRT model has (that there is something like the Item Characteristic Curve)
- but only when the representation of the latent trait is correct!
- Mantel-Haenszel less prone to this error

Comparison

- possibilities are:
  - using only items that have in previous studies been shown to have no DIF as the representation of the latent trait
  - using only items that have been shown in a first round of analyses on this data set that they do not show any DIF to represent the latent trait („purification“)
- both are possible in R

Logistic Regression to detect DIF

- new result
- with purification

old result

Logistic regression DIF statistic:

	Stat.	P-value	
V1	2.1052	0.1468	
V2	13.6294	0.0002	***
V3	0.2270	0.6337	
V4	22.3568	0.0000	***
V5	0.8874	0.3462	
V6	89.7839	0.0000	***
V7	0.4740	0.4911	
V8	0.3492	0.5545	
V9	0.4654	0.4951	
V10	39.7758	0.0000	***
V11	1.1193	0.2901	
V12	13.4808	0.0002	***
V13	0.0089	0.9248	
V14	1.8489	0.1739	
V15	29.8718	0.0000	***

ERSITY OF
IBRIDGE
Psychometri

Logistic regression DIF statistic:

	Stat.	P-value	
V1	0.4684	0.4937	
V2	9.1764	0.0025	**
V3	2.1723	0.1405	
V4	36.5388	0.0000	***
V5	0.0996	0.7523	
V6	84.0787	0.0000	***
V7	0.4577	0.4987	
V8	0.9172	0.3382	
V9	4.8994	0.0269	*
V10	31.8524	0.0000	***
V11	6.6677	0.0098	**
V12	32.4617	0.0000	***
V13	2.4271	0.1193	
V14	6.0825	0.0137	*
V15	23.5556	0.0000	***

MH polytomous

- only implemented by the mantelhaen.test() command
  - here the score has to be defined
  - and it has to be done for every item by hand…

Logistic Regression polytomous

- implemented in the „lordif“ package
- again Big5 data (recoded version from Anna‘s folder with the DIFAS practical)

DIF polytomous

- Results Big5 in DIFAS
- items 14, 19 and 24

DIF STATISTICS: POLYTOMOUS ITEMS

Name	Mantel	L-A LOR	LOR SE	LOR Z	COX'S B	COX SE	COX Z
Var 14	8.418	0.583	0.206	2.83	0.439	0.1513	2.902
Var 15	0.319	-0.107	0.189	-0.566	-0.066	0.1177	-0.561
Var 16	0.154	-0.085	0.215	-0.395	-0.062	0.1577	-0.393
Var 17	0.005	-0.013	0.192	-0.068	-0.009	0.1288	-0.07
Var 18	0.367	0.118	0.195	0.605	0.075	0.1233	0.608
Var 19	14.052	0.74	0.196	3.776	0.472	0.1258	3.752
Var 20	0.009	-0.018	0.198	-0.091	-0.012	0.1321	-0.091
Var 21	0.236	-0.109	0.222	-0.491	-0.083	0.1705	-0.487
Var 22	0.02	-0.029	0.206	-0.141	-0.019	0.1347	-0.141
Var 23	2.093	-0.304	0.206	-1.476	-0.215	0.1488	-1.445
Var 24	10.911	-0.631	0.191	-3.304	-0.378	0.1143	-3.307
Var 25	0.181	0.089	0.209	0.426	0.062	0.1464	0.423

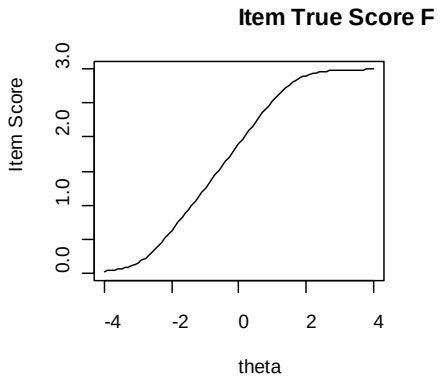
Reference Value = 1, Focal Value = 2

Logistic Regression polytomous

- this package contains the command lordif()
- estimates the latent trait via Graded Response Model (from ltm)
- conditions in this case on estimated thetas and not on the score
- purification is always performed

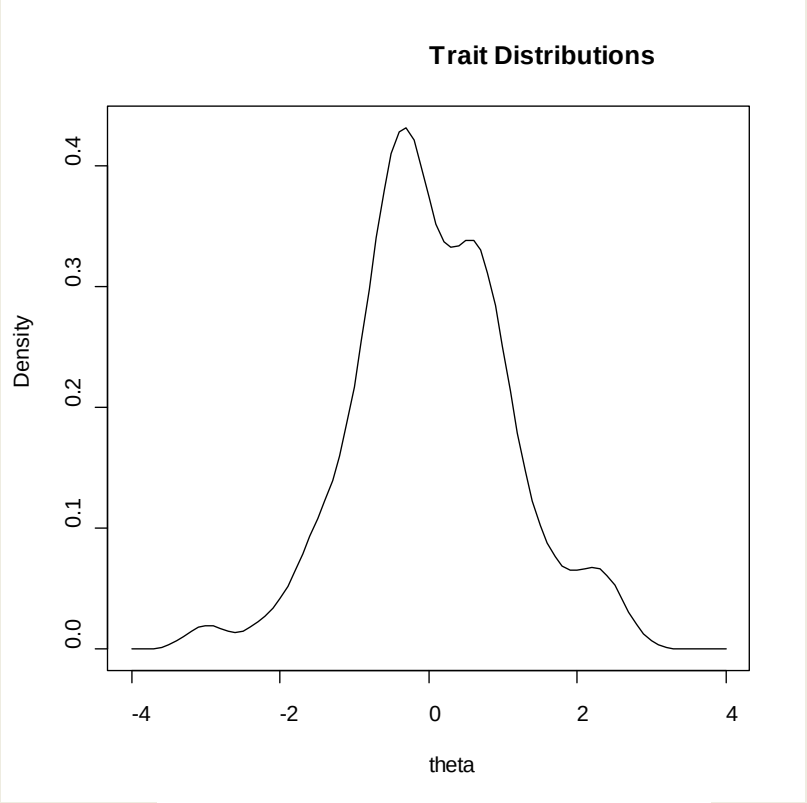
Logistic Regression polytomous

- items 14 and 19 are detected to show DIF



Logistic Regression polytomous

- distributions of our latent trait in the two populations

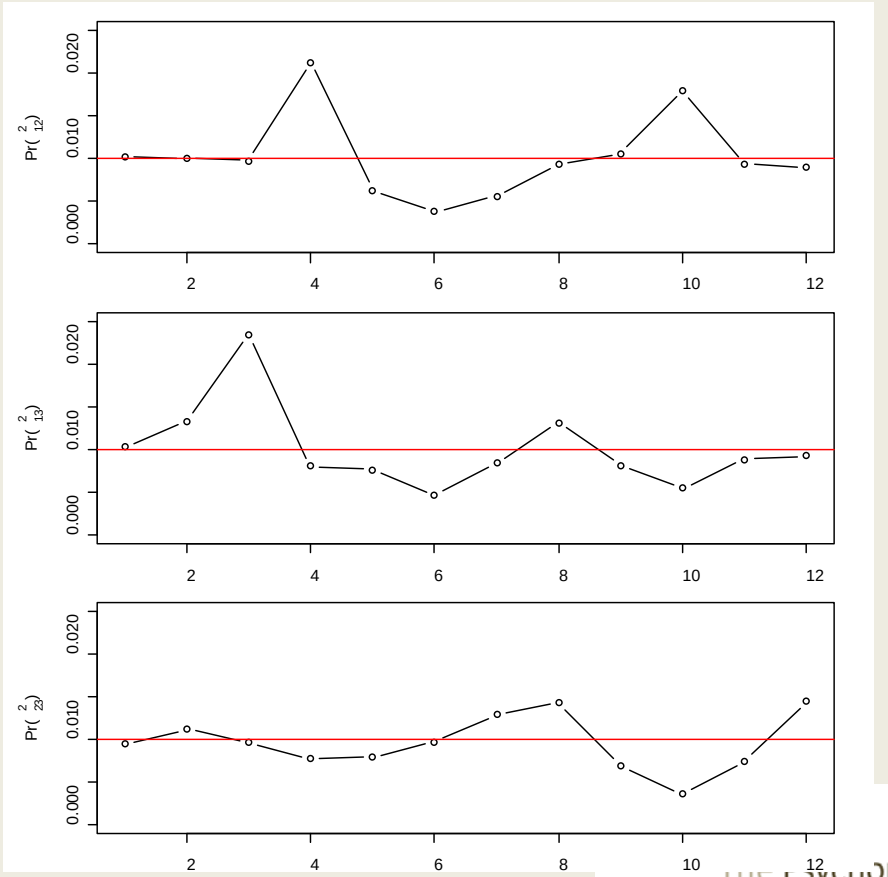


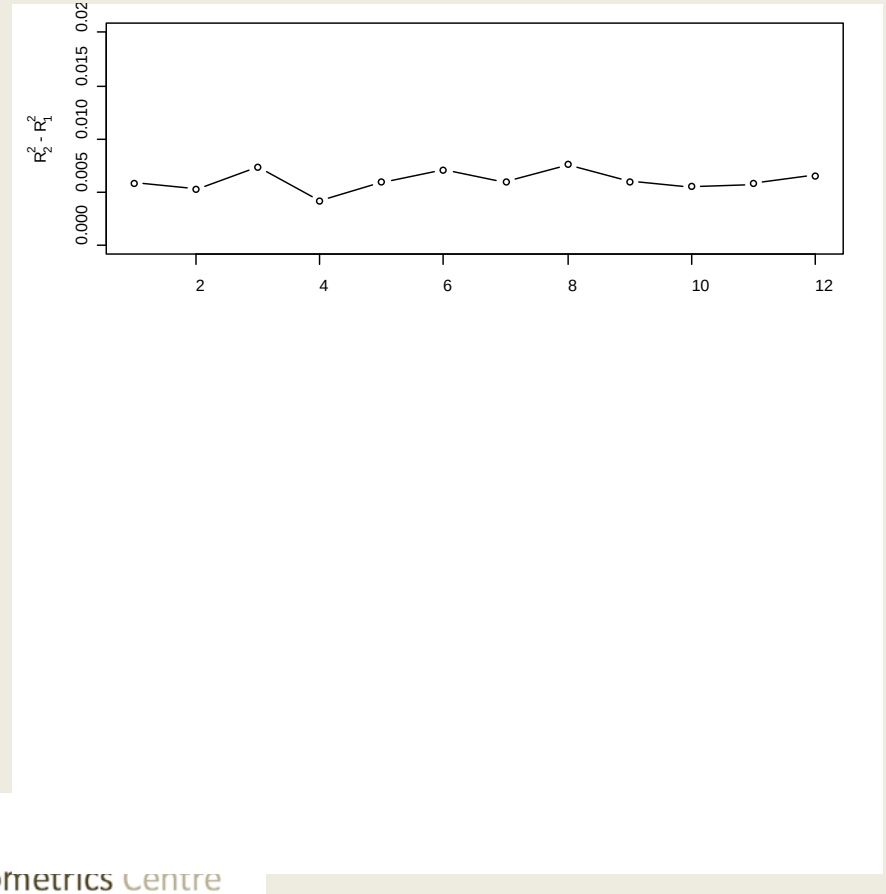
Logistic Regression polytomous

- lordif has also a montecarlo command that can be used to validate significance levels
- it simulates datasets with the item parameters – assuming that there was no DIF present
- performs the same tests on these data sets and so a comparison of the resulting test statistics with the empirical from the current data set is possible

Logistic Regression polytomous

- simulated p-values and differnces





Differential Test Functioning

- not implemented in R (yet?)