

การวิเคราะห์แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน ด้วยโมเดลโลจิสติก

ผศ.ดร. พงษ์จิต อินทสุวรรณ



สถาบันวิจัยพฤติกรรมศาสตร์
มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร
มิถุนายน 2534

รายงานการวิจัยฉบับที่ 45

การวิเคราะห์แบบทดสอบวัดผลสัมฤทธิ์
ทางการเรียน ด้วยโมเดลเลจิสติก

ผศ.ดร. พงจิต อินทสุวรรณ

สถาบันวิจัยพฤติกรรมศาสตร์

มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร

มิถุนายน 2534

คำนำ

งานวิจัยฉบับที่ 45 ของสถาบันวิจัยพฤติกรรมศาสตร์ เรื่องการวิเคราะห์แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนด้วยโมเดลโลจิสติก มีจุดมุ่งหมายสำคัญหลายประการ คือ (1) เพื่อแสดงขั้นตอน วิธีการ และผลการวิเคราะห์รายชื่อของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน ด้วยการใช้โมเดลโลจิสติก (2) เพื่อเปรียบเทียบผลการวิเคราะห์รายชื่อของแบบทดสอบที่ใช้โมเดลโลจิสติกกับที่ใช้โมเดลดั้งเดิม และ (3) เพื่อตรวจสอบความสอดคล้องและความขัดแย้งในผลการวิเคราะห์รายชื่อที่ใช้โมเดลโลจิสติกกับที่ใช้โมเดลดั้งเดิม

งานวิจัยค้นคว้าทางวิธีการขั้นนี้จะเป็นประโยชน์อย่างยิ่งต่อครูอาจารย์ นักการศึกษา นิสิตนักศึกษา ตลอดจนผู้สนใจในโมเดลใหม่ ๆ ทางด้านวัดผลเพื่อที่จะได้นำวิธีการนี้ไปใช้ในการคัดเลือกข้อสอบให้เหมาะสมยิ่งขึ้น

สถาบันใดที่ขอขอบคุณ ผู้ช่วยศาสตราจารย์ ดร.พจนจิต อินทสุวรรณ ที่ได้ให้ความมุ่งมั่นอุตสาหะในการศึกษางานวิจัยขั้นนี้จนประสบผลสำเร็จ ยังให้เกิดผลงานที่มีคุณค่าและประโยชน์ในแง่วิชาการอย่างยิ่ง

จรจ สุวรรณทัต

(ศาสตราจารย์ ดร.จรจ สุวรรณทัต)

ผู้อำนวยการสถาบันวิจัยพฤติกรรมศาสตร์

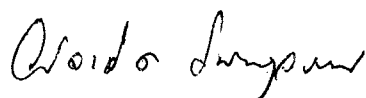
มิถุนายน 2534

กิตติกรรมประกาศ

งานวิจัยเรื่องนี้เป็นไปได้ด้วยได้รับความร่วมมือจากสำนักงานการศึกษารุงเทพมหานครอนุญาตให้ใช้ผลการสอบวัดผลสัมฤทธิ์ปลายปี วิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2529 ของกลุ่มโรงเรียนในเขตบางกะปิ เป็นข้อมูลเพื่อการวิเคราะห์ ผู้วิจัยซาบซึ้งในความร่วมมือครั้งนี้เป็นอย่างยิ่ง และขอขอบคุณคุณวิชัย เอียดบัว มหาบัณฑิตจากสถาบันวิจัยพฤติกรรมศาสตร์ ที่ได้ประสานงานในการขอใช้ข้อมูลชุดนี้

เนื่องจากแบบทดสอบนี้เป็นความลับของทางราชการ ผู้วิจัยจึงไม่นำมาลงไว้ในรายงานฉบับนี้

ผู้วิจัยขอขอบคุณผู้ที่มีส่วนช่วยเหลือในการวิเคราะห์ข้อมูลด้วยเครื่องคอมพิวเตอร์ ดังรายนามต่อไปนี้ ดร. ชูศักดิ์ ชัมภลิติต แห่งกระทรวงศึกษาธิการ คุณภาณี บุญาค แห่งสถาบันเทคโนโลยีแห่งเอเชีย และผู้ช่วยศาสตราจารย์ลัดดาวัลย์ เกษมเนตร แห่งสถาบันวิจัยพฤติกรรมศาสตร์



(ผศ.ดร. พจจิต อินทสุวรรณ)

20 มิถุนายน 2534

สารบัญ

หน้า

รายชื่อตาราง.....

รายชื่อภาพ

บทที่ 1 บทนำ.....	1
วัตถุประสงค์และคำถามของการวิจัย.....	4
ทฤษฎีการตอบข้อคำถาม.....	5
ข้อตกลงเบื้องต้นที่สำคัญ.....	10
วิธีการในการตรวจสอบความเป็นมติเดียวของข้อคำถาม.....	13
การประมาณค่าพารามิเตอร์.....	17
การประมาณค่าพารามิเตอร์เมื่อผู้สอบเว้นบางข้อ.....	21
วิธีการประมาณค่าพารามิเตอร์ในโปรแกรมคอมพิวเตอร์ LOGIST	23
วิธีการประมาณค่าพารามิเตอร์สำหรับโมเดลของราช.....	25

บทที่ 2 ระเบียบวิธี.....	28
แบบทดสอบ.....	28
กลุ่มตัวอย่าง.....	28
วิธีการตรวจสอบความเป็นมติเดียว.....	28
การประมาณค่าพารามิเตอร์.....	30
การหาความสัมพันธ์ระหว่างค่าประมาณความชุกจากโมเดลทั้งสาม	32

	หน้า
บทที่ 3 ผลการวิเคราะห์ข้อมูล.....	33
ผลการวิเคราะห์เนื้อหาและพฤติกรรมที่แบบทดสอบฉบับนี้วัด.....	33
ผลการตรวจสอบความเป็นมิติเดียวของคะแนนจากแบบทดสอบ....	33
ข้อมูลพื้นฐานของคะแนนรวมจากแบบทดสอบ.....	36
ผลการประมาณค่าพารามิเตอร์ภายใต้โมเดลโลจิสติกที่มีพารามิเตอร์	
3 ตัว (โมเดล 3PL).....	36
ผลการประมาณค่าพารามิเตอร์ในโมเดลของราช (โมเดล 1PL)..	44
ผลการวิเคราะห์รายข้อตามโมเดลดั้งเดิม.....	53
ความสัมพันธ์ระหว่างค่าความยากที่ประมาณจากโมเดลโลจิสติกและ	
โมเดลดั้งเดิม.....	57
บทที่ 4 สรุปและอภิปรายผล.....	59
ผลโดยสรุป.....	59
อภิปรายผล.....	61
ความสัมพันธ์ระหว่างค่าประมาณความยากในโมเดลทั้งสาม..	61
ปัญหาของข้อคำถามบางข้อ.....	63
สรุปการอภิปรายผล.....	69
ข้อเสนอแนะสำหรับการปฏิบัติ.....	71
เอกสารอ้างอิง.....	73

	หน้า
บทที่ 3 ผลการวิเคราะห์ข้อมูล.....	33
ผลการวิเคราะห์เนื้อหาและพฤติกรรมที่แบบทดสอบฉบับนี้วัด.....	33
ผลการตรวจสอบความเป็นมิติเดียวของคะแนนจากแบบทดสอบ....	33
ข้อมูลพื้นฐานของคะแนนรวมจากแบบทดสอบ.....	36
ผลการประมาณค่าพารามิเตอร์ภายใต้โมเดลโลจิสติกที่มีพารามิเตอร์	
3 ตัว (โมเดล 3PL).....	36
ผลการประมาณค่าพารามิเตอร์ในโมเดลของราช (โมเดล 1PL)..	44
ผลการวิเคราะห์รายข้อตามโมเดลดั้งเดิม.....	53
ความสัมพันธ์ระหว่างค่าความยากที่ประมาณจากโมเดลโลจิสติกและ	
โมเดลดั้งเดิม.....	57
บทที่ 4 สรุปและอภิปรายผล.....	59
ผลโดยสรุป.....	59
อภิปรายผล.....	61
ความสัมพันธ์ระหว่างค่าประมาณความยากในโมเดลทั้งสาม..	61
ปัญหาของข้อคำถามบางข้อ.....	63
สรุปการอภิปรายผล.....	69
ข้อเสนอแนะสำหรับการปฏิบัติ.....	71
เอกสารอ้างอิง.....	73

ภาคผนวก:

ภาคผนวก ก	แสดงหมายเลขข้อในแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ชั้นประถมศึกษาปีที่ 6 จำแนกตามเนื้อหา และพฤติกรรม.....	81
ภาคผนวก ข	โต้แย้งแสดงลักษณะของข้อคำถาม 59 ข้อ ภาษาใต้โมเดล 3PL.....	84
ภาคผนวก ค	โต้แย้งแสดงลักษณะของข้อคำถามภาษาใต้โมเดล 1PL	115
ภาคผนวก ง	แสดงค่า p ทั้ง 6 ค่าในช่วงคะแนน 6 ช่วงสำหรับแต่ละข้อ.....	147
ภาคผนวก จ	แสดงรายละเอียดผลการวิเคราะห์รายข้อตามแบบดั้งเดิม.....	152
ภาคผนวก ฉ	แสดงอันดับของข้อคำถาม 59 ข้อ ในแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ตามค่าความยากที่ประมาณภาษาใต้โมเดล 3PL โมเดล 1PL และโมเดลดั้งเดิม.....	168

รายชื้อตาราง

ตาราง

หน้า

1	แสดงจำนวนนักเรียนในกลุ่มตัวอย่างจำแนกตามโรงเรียน.....	29
2	แสดงจำนวนข้อในแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิทยาศาสตร์ ชั้นประถมศึกษาปีที่ 6 จำแนกตามเนื้อหาและพฤติกรรมต่าง ๆ ที่วัด.....	34
3	แสดงการแจกแจงความถี่ของคะแนนรวมจากแบบทดสอบวัดผลสัมฤทธิ์ วิทยาศาสตร์ที่มี 60 ข้อ ของนักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 1,502 คน.....	37
4	แสดงค่าประมาณพารามิเตอร์อำนาจจำแนก (a) ความชาก (b) และ การเดา (c) ของข้อคำถาม 59 ข้อ ในแบบทดสอบวัดผลสัมฤทธิ์ วิทยาศาสตร์ จากผลการตอบของนักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 1,502 คน.....	40
5	แสดงค่าประมาณพารามิเตอร์ความชาก (b) ความคลาดเคลื่อนมาตรฐาน (Std error) และการทดสอบทีของความพอเหมาะ (Fit t-test) ของข้อคำถาม จำนวน 60 ข้อในแบบทดสอบวัดผลสัมฤทธิ์วิทยาศาสตร์ ของนักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 1,485 คน.....	45
6	แสดงค่าประมาณพารามิเตอร์ความชาก (b) ความคลาดเคลื่อนมาตรฐาน (Std error) อำนาจจำแนก (Disc index) และการทดสอบทีของ ความพอเหมาะ (Fit t-test) ของข้อคำถามจำนวน 59 ข้อ ในแบบ ทดสอบวัดผลสัมฤทธิ์วิทยาศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 1,485 คน.....	49

7 แสดงดัชนีความง่าย (p) และอำนาจจำแนก (r_{1u}) ของข้อคำถาม 60 ข้อ ในแบบทดสอบวัดผลสัมฤทธิ์ วิชาคณิตศาสตร์ของนักเรียนชั้น ประถมศึกษาปีที่ 6 จำนวน 1,502 คน.....	54
8 แสดงหมายเลขข้อ (พร้อมทั้งตัวเลือก) ที่มีลักษณะไม่สอดคล้องกับ หลักของข้อคำถามที่ดี.....	57
9 แสดงค่าสัมประสิทธิ์สหสัมพันธ์แบบอันดับ (สเปียร์แมน r_s) ระหว่าง ค่าประมาณความยากจากโมเดล 3PL 1PL และดั้งเดิม ของข้อ คำถาม 59 ข้อ จากแบบทดสอบวัดผลสัมฤทธิ์วิชาคณิตศาสตร์.....	58

รายชื่อภาพ

ภาพ	หน้า
1 โค้งแสดงลักษณะของข้อคำถาม.....	10
2 แสดงค่าไอเกินหาค่าแรกและค่าร้อยละของความแปรปรวนที่แต่ละ แพคเตอร์อธิบายคะแนนรวมได้.....	35
3 แสดงรูปหลายเหลี่ยมแห่งความถี่ของคะแนนรวมจากแบบทดสอบวัดผล สัมฤทธิ์วิชาคณิตศาสตร์.....	39

บทคัดย่อ

การวิเคราะห์แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนด้วยโมเดลโลจิสติก

จุดมุ่งหมายในการศึกษาค้างนี้มี 3 ประการคือ ประการแรก เพื่อแสดงขั้นตอนวิธีการ และผลการวิเคราะห์รายชื่อจากการใช้โมเดลโลจิสติก ประการที่สอง เพื่อเปรียบเทียบผลการวิเคราะห์รายชื่อโดยใช้โมเดลโลจิสติกกับโมเดลดั้งเดิม และประการที่สามเพื่อตรวจสอบความสอดคล้องและความขัดแย้งระหว่างผลการวิเคราะห์รายชื่อจากการใช้โมเดลโลจิสติกกับผลจากการใช้โมเดลดั้งเดิม

ข้อมูลที่ใช้ในการวิเคราะห์ครั้งนี้เป็นผลการสอบวิชาคณิตศาสตร์ประจำภาคปลาย ปีการศึกษา 2529 ของนักเรียนชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2529 จำนวน 1,502 คน ในกลุ่มโรงเรียนเขตบางกะปิ กรุงเทพมหานคร แบบทดสอบวัดผลสัมฤทธิ์ฉบับนี้เป็นแบบเลือกตอบมีจำนวน 60 ข้อ

ผลการตรวจสอบความเป็นมิติเดียวแสดงว่าคะแนนรวมมีมิติเดียว

การประมาณค่าพารามิเตอร์ของข้อภายใต้โมเดลโลจิสติกที่มีพารามิเตอร์ 3 ตัว (โมเดล 3PL) (โดยตัดข้อที่ 14 ออก) ปรากฏผลดังนี้ ความยากมีค่าจาก -3.7490 ถึง 3.9110 อำนาจจำแนกมีค่าจาก .1471 ถึง 2.0000 และแอซิมโทตล่างหรือการเดามีค่าจาก .0418 ถึง .3966

ข้อที่มีอำนาจจำแนกต่ำกว่า .3000 มี 4 ข้อ คือ ข้อ 15, 30, 41, และ 60

ค่าประมาณพารามิเตอร์ของข้อ (ความยาก) ภายใต้โมเดลของราช (โมเดล 1PL) (ตัดข้อที่ 14 ออก) มีช่วงจาก -2.68 ถึง 1.90 การทดสอบว่าโมเดลเหมาะสมกับข้อมูลหรือไม่โดยใช้สถิติที่ปรากฏว่ามี 4 ข้อ ที่ไม่เหมาะสมคือข้อ 30, 34, 45, และ 60

สำหรับการวิเคราะห์รายชื่อโดยใช้โมเดลดั้งเดิม (ใช้ทั้งหมด 60 ข้อ) ปรากฏผลดังนี้ ค่าประมาณความง่าย (สัดส่วนการตอบถูก) มีช่วงจาก .1853 ถึง .9401 ข้อที่ค่าความง่ายสูงกว่า .9000 มี 3 ข้อ คือข้อ 1, 3, และ 17 ค่าประมาณอำนาจจำแนก (สหสัมพันธ์พ้อยท์ไบซีเรียลระหว่างข้อและคะแนนรวมทั้งฉบับที่ชกเว้นข้อนั้น) มีช่วงจาก -.0891 ถึง .4170 ข้อที่ 14 มีค่าอำนาจจำแนกต่ำสุด และเป็นค่าลบ ข้อที่มีค่าอำนาจจำแนกต่ำกว่า .10 มี 5 ข้อ คือ ข้อ 14, 30, 34, 58, และ 60

การวิเคราะห์สัดส่วนการตอบตัวเลือกต่าง ๆ เพื่อดูว่าการเลือกเป็นไปตามหลักการที่เหมาะสมหรือไม่ ผลแสดงว่าการตอบตัวเลือกที่ไม่เป็นไปตามหลักการที่เหมาะสมใน 1 ตัวเลือกมี 18 ข้อและใน 2 ตัวเลือกมี 3 ข้อ

อันดับความยากของข้อภายใต้โมเดลทั้งสามมีความสอดคล้องกันสูงมาก โดยเฉพาะอย่างยิ่งระหว่างโมเดล 1PL กับโมเดลดั้งเดิม

โมเดล 3PL ปฏิเสธข้อคำถามเป็นจำนวนน้อยกว่าโมเดล 1PL โมเดลดั้งเดิมให้ผลที่ขึ้นอยู่กับการสุ่ม แต่โมเดลดั้งเดิมนี้ก็ให้ผลที่จะเป็นส่วนประกอบกับผลจากโมเดลโลจิสติกได้ดี ในแง่ของการพิจารณาปรับปรุงตัวเลือก ดังนั้นในขั้นตอนของการพัฒนาแบบทดสอบ นั่นคือ วิเคราะห์ข้อคำถามเพื่อปรับปรุงแก้ไขข้อคำถามและตัวเลือก จึงควรใช้โมเดลดั้งเดิมประกอบกับโมเดลโลจิสติก เมื่อได้แบบทดสอบฉบับสมบูรณ์แล้วควรประมาณค่าพารามิเตอร์ของข้อภายใต้โมเดลโลจิสติก

ABSTRACT

An analysis of achievement test using logistic models

This study had three main objectives. First, to demonstrate the procedures of item analysis, using logistic models, including the interpretation of the results. Second, to compare the results of item analyses using logistic models and the classical model. Third, to examine the agreement and the contradiction of the results from the logistic models and the classical model.

The data used in this study were responses of 1502 Prathom Suksa VI students to the 60-item mathematics achievement test. The testing, a regular school examination, was taken place at the end of the second semester of the academic year 1986. These students were from schools in Bangkapi area in Bangkok.

The preliminary analyses indicated that the test scores were unidimensional.

Under the three parameter logistic model -- 3PL model, the parameter estimation results were as follows: difficulty estimates ranged from -3.7490 to 3.9110, discrimination estimates ranged from .1471 to 2.000, and the lower asymptote or guessing estimates ranged from .0418 to .3966. Item no. 14 caused the incompleted run, therefore it was excluded from the

estimation. Four items had discrimination estimates less than .3000, these were item nos. 15, 30, 41, and 60.

Under the Rasch model -- 1PL model, the difficulty parameter estimates of 59 items were between -2.68 and 1.90. The fit analysis showed that four items did not fit the model. They were item nos. 30, 34, 45, and 60.

All sixty items were included in the item analysis under the classical model. The easyness parameter estimates -- proportion correct, ranged from .1853 to .9401. Three items had easyness indices above .9000, these were item nos. 1, 3, and 17. The discrimination estimates --point biserial correlation between item i and total test excluding item i , ranged from -.0891 to .4170. Item no. 14 had the lowest discrimination, and it was negative. Five items had discrimination indices lower than .10, these were item nos. 14, 30, 34, 58, and 60.

Further analyses were done to examine whether or not the choices chosen by the examinees, among 4 alternatives for each item, were in accord with the examiner's expectations. The results showed that from 18 items out of 60, only one alternative was found inappropriate for each item, and 2 alternatives from another 3 items were of the same problem.

Ranks of item difficulties estimated from the three different models were in high agreement, especially between the 1PL model and the classical model.

The 3PL model rejected less number of items than the 1PL model did. Eventhough the classical model gave results that dependent upon the group ability, it gave useful information for the alternative improvement. Therefore, at the stage of test development, the classical model should be used in couple with the logistic model. For the final test forms, the item parameters should be estimated under the logistic model.

บทที่ 1

บทนำ

แบบทดสอบหรือข้อสอบ (test) ประกอบด้วยงานหรือคำถามชุดหนึ่ง ซึ่งเมื่อนำเสนอต่อผู้สอบภายใต้สถานการณ์ที่เป็นมาตรฐาน จะกระตุ้นพฤติกรรมบางอย่างออกมา และมีการตรวจให้คะแนน ซึ่งคะแนนนี้มีคุณสมบัติตามหลักการวัดทางจิตวิทยา เช่น มีความเชื่อมั่นและความเที่ยงตรงสูง (Davis, 1974, p.2) พฤติกรรมหรือคุณลักษณะที่ผู้สร้างข้อสอบตั้งใจจะวัดเป็นสิ่งที่แฝงอยู่ในตัวบุคคล ไม่สามารถสังเกตได้ จึงต้องประเมินโดยอาศัยผลการตอบข้อคำถาม ตัวอย่างของคุณลักษณะดังกล่าว เช่น ความถนัดทางการเรียน ผลสัมฤทธิ์ทางการเรียน ความสนใจ ทักษะคิด และลักษณะนิสัย เป็นต้น ถ้าคำนึงถึงคุณลักษณะที่จะวัดเป็นเกณฑ์แบบทดสอบสามารถแบ่งออกได้เป็นสองกลุ่ม (Cronbach, 1970, p.35-42) กลุ่มแรกเป็นการวัดความสามารถสูงสุด และกลุ่มที่สองเป็นการวัดพฤติกรรมตามปกติ ซึ่งมักเรียกชื่อแตกต่างไปจากแบบทดสอบ เช่น แบบสำรวจ แบบทดสอบในกลุ่มแรก รวมแบบทดสอบที่วัดความสามารถทางสมองทั่วไปหรือสติปัญญา ความสามารถเฉพาะ และความเชี่ยวชาญในสิ่งที่ได้ฝึกฝนมาแล้วซึ่งรวมทั้งผลสัมฤทธิ์ทางการเรียนด้วย การศึกษาค้นคว้านี้ใช้คะแนนจากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ระดับชั้นประถมศึกษาปีที่ 6 ของกลุ่มโรงเรียนกลุ่มหนึ่งในสังกัดกรุงเทพมหานคร ปีการศึกษา 2529 คะแนนจากแบบทดสอบที่วัดความสามารถสมองได้รับความสนใจอย่างสูง และมีการพัฒนาอย่างต่อเนื่องเกี่ยวกับโมเดลทางคณิตศาสตร์ที่แทนคะแนน เพื่อแก้ปัญหาเกี่ยวกับความคลาดเคลื่อนในการวัด (Gulliksen, 1950; and Lord and Novick, 1968) เนื่องจากคุณลักษณะที่เราสนใจวัดนั้นเป็นสิ่งที่ไม่สามารถวัดได้โดยตรง ต้องศึกษาทางอ้อมโดยวิธีวัดปริมาณอื่น ๆ เช่น เราไม่สามารถวัดความสามารถทางคณิตศาสตร์ของ

เด็กได้เราเพียงแต่วัดผลการตอบข้อคำถามต่างๆ ทางคณิตศาสตร์ของเขา

แบบทดสอบฉบับหนึ่ง ๆ ประกอบด้วยข้อคำถามหลายข้อ ผู้สร้างแบบทดสอบเขียนข้อคำถามขึ้นเป็นจำนวนมากกว่าจำนวนข้อที่ต้องการ ทดลองสอบ แล้ววิเคราะห์รายชื่อเพื่อให้ได้ข้อมูลสำหรับประกอบการพิจารณาเพื่อปรับปรุงหรือคัดเลือกข้อคำถามแบบทดสอบฉบับที่ได้ย่อมจะให้คะแนนที่มีคุณภาพ นั่นคือมีความเชื่อมั่น และเที่ยงตรงสูง

การวิเคราะห์รายชื่อกระทำได้ทั้งในด้านการบรรยาย เช่น เนื้อหา และรูปแบบ และในด้านปริมาณ คือ การคำนวณค่าสถิติประจำข้อ (Anastasi, 1976) ในทางปฏิบัติควรใช้ทั้งสองแนวประกอบกัน นั่นคือ การคัดเลือกข้อควรคำนึงถึงทั้งค่าสถิติ และ เนื้อหารวมทั้งพิจารณาจุดมุ่งหมายในการวัดด้วย การวิเคราะห์รายชื่อในด้านปริมาณมีหลักการและวิธีการที่เป็นปรนัยโดยอาศัยสูตรทางสถิติ และตำราเกี่ยวกับการทดสอบให้มีความสำคัญมาก การศึกษาครั้งนี้เป็นเรื่องของการวิเคราะห์รายชื่อในด้านปริมาณ

แนวความคิดเกี่ยวกับการวิเคราะห์รายชื่อเริ่มจากงานของ บิเน็ต (Binet) ในปี 1916 (Anastasi, 1976, p.230-245) ในการสร้างแบบทดสอบวัดสติปัญญาของคน บิเน็ตเชื่อว่าสติปัญญาหรือความสามารถสมองของคนเพิ่มขึ้นตามอายุ เขาตรวจให้คะแนนข้อคำถามแต่ละข้อเป็นแบบผ่าน-ไม่ผ่าน ตามเกณฑ์ที่กำหนดไว้ว่าสำหรับอายุระดับนั้น ๆ ควรสามารถทำได้มากเพียงใด แล้วพล็อตจุดที่แทนความสัมพันธ์ ระหว่างสัดส่วนการตอบถูกสำหรับแต่ละข้อกับระดับอายุของผู้สอบ เกณฑ์ในการคัดเลือกข้อคำถามไว้ในแบบทดสอบวัดสติปัญญา คือ พิจารณาคัดเลือกข้อที่มีสัดส่วนการตอบถูกเพิ่มขึ้นตามอายุ แบบทดสอบวัดสติปัญญาของบิเน็ตจึงนิยมใช้กันทั่วไปกับเด็กมากกว่ากับผู้ใหญ่

หลักการพื้นฐานในการวิเคราะห์รายชื่อ คือ โมเดลในการวัดทางจิตวิทยาที่ใช้กันมานานซึ่งอธิบายลักษณะของข้อคำถามด้วยความยาก (item difficulty)

และ ความเที่ยงตรง (item validity) (Guilford, 1954, p.417-456; Henrysson, 1976, p.138-143; and Anastasi, 1976, p.198-226)

ความยากของข้อนิยามในเทอมของเบอร์เช็นด์ของคนสอบที่ตอบข้อนั้น ได้ถูก (ใช้สัญลักษณ์ p) ข้อที่มีคนตอบถูกน้อยกว่าหรือไม่มีคนตอบถูกเลยย่อมไม่มีประโยชน์ในแง่ของการกระจายของคะแนน ข้อที่มีคนตอบถูกแทบทั้งหมดหรือทั้งหมดก็เช่นเดียวกัน ดังนั้นจึงนิยามคัดเลือกข้อที่มีค่าของ p ปานกลาง โดยมีค่าเฉลี่ยเป็น .5 ค่า p นี้อยู่ในมาตราอันดับ (ordinal scale) จึงมีการแปลง p ให้เป็น Δ (อ่านว่า เดลต้า (delta)) ซึ่งอยู่ในมาตราช่วงเท่ากัน (interval scale)

นิยามของความยากที่แสดงด้วยค่า p นั้น ทำให้เกิดความสับสนในการตีความ นั่นคือค่า p ยิ่งสูงข้อนั้นยิ่งยากน้อยลง (ง่ายขึ้น) ดังนั้นในงานนี้จึงเรียก p ว่าเป็นความง่าย

กระบวนการในการกำหนดความเที่ยงตรงของข้อแบ่งออกเป็น 4 กลุ่ม กลุ่มแรกใช้ทฤษฎีที่ว่าความน่าจะเป็นในการตอบข้อหนึ่งได้ถูกเป็นฟังก์ชันแบบปกติสะสม (ogive function) ของความสามารถ กลุ่มที่สองเน้นความสามารถของข้อในการจำแนกกลุ่มสูง และกลุ่มต่ำ กลุ่มที่สามเน้นการหาความสัมพันธ์ระหว่างข้อกับเกณฑ์ สำหรับเกณฑ์ที่ใช้ในกลุ่มที่สอง และสามนั้นรวมทั้งเกณฑ์ภายในหรือคะแนนรวมทั้งฉบับด้วย วิธีการของกลุ่มที่สามนิยมใช้กันมาก และสัมพันธ์สหสัมพันธ์ที่เลือกให้ได้มี 4 ชนิด คือ สหสัมพันธ์ไบซีเรียล สหสัมพันธ์พัสคัล-ไบซีเรียล สหสัมพันธ์เทตเรลลิด และสัมพันธ์สหสัมพันธ์ที่เป็นการวิเคราะห์ความแปรปรวน

ในช่วงเวลาต่อมา (ประมาณปี ค.ศ.1960 เป็นต้นมา) มีความก้าวหน้าที่สำคัญในเรื่อง การวิเคราะห์รายข้อโดยใช้โมเดลโค้งแสดงลักษณะข้อคำถาม

(item characteristic curve model--icc model) เค็งนี้อาจใช้ฟังก์ชัน เค็งปกติสะสม (normal ogive) หรือโลจิสติก (logistic) แต่ฟังก์ชันโลจิสติก ได้รับความนิยมนมากกว่าเนื่องจากท้อได้เปรียบในด้านของรูปแบบฟังก์ชันและคุณสมบัติ เริงคณิตศาสตร์ โมเดลนี้อยู่ภายใต้ทฤษฎีคุณลักษณะแฝง (latent trait theory) หรือทฤษฎีการตอบท้อคำถาม (item response theory) (ดู Lord and Novick, 1968; Lord, 1980; Hambleton and Swaminathan, 1985 และ ผจจจิต ลินทสุวรรณ, 2528) ทฤษฎีนี้มีพื้นฐานทางสถิติขั้นสูง มีวิธีการประมาณค่าพารามิเตอร์ ของผู้สอบ (ความสามารถ) และ พารามิเตอร์ของข้อที่ซับซ้อนซึ่งต้องอาศัยคอมพิวเตอร์ ในการคำนวณ อย่างไรก็ตามการให้โมเดลใหม่นี้สำหรับการวิเคราะห์รายข้อโดยเฉพาะใน ประเทศไทยคงจำกัต้องอยู่ในกลุ่มนักวิชาการ และหน่วยงานของชาติที่ทำหน้าที่เกี่ยวกับการทดสอบ

วัตถุประสงค์และคำถามของการวิจัย

นักวิชาการยอมรับกันว่ามีช่องว่างระหว่างพัฒนาการในทางทฤษฎีกับ การใช้ในสถานการณ์จริง สำหรับผู้ปฏิบัติ นอกจากจะมีปัญหาเกี่ยวกับความสั้เปลี่ยนแปลง ในการคำนวณแล้ว ปัญหาที่สำคัญกว่าก็คือ การทำความเข้าใจ เกี่ยวกับโครงสร้าง ของทฤษฎี วิธัค่าเนิการและการตีความผล การศึกษาคั้งนี้จึงมีวัตถุประสงค์

3 ประการคือ

1. เพื่อให้สังกัปลั้ของทฤษฎีการตอบท้อคำถาม
2. เพื่อแสดงขั้นตอน วิธีการ และผล (รวมทั้งการตีความผล) การ วิเคราะห์รายข้อโดยให้โมเดลโลจิสติก พร้อมด้วผลการวิเคราะห์ตามโมเดลดั้งเดิม
3. เพื่อตรวจสอบว่าโมเดลโลจิสติกที่ใช้ในการศึกษาคั้งนี้ คือ โมเดล

โลจิสติกที่มีพารามิเตอร์ 3 ตัว และ 1 ตัว กับโมเดลดั้งเดิม จะให้ผลที่สอดคล้องกัน
เพียงใด และ มีส่วนที่วาทสนับสนันกันเพียงใดในแง่ของการเข้าใจข้อคำถามที่วิเคราะห์
ข้อมูลที่ใช้เป็นผลการตอบข้อคำถาม ในแบบทดสอบวัดผลสัมฤทธิ์ วิชาเลขคณิต
ของนักเรียนชั้นประถมศึกษาปีที่ 6 ในโรงเรียนบางโรงในกรุงเทพมหานคร

ทฤษฎีการตอบข้อคำถาม

ทฤษฎีการตอบข้อคำถาม (item response theory) แสดง
ความสัมพันธ์ระหว่างการตอบข้อคำถามข้อหนึ่ง ๆ กับ คุณลักษณะที่ข้อสอบเจอนั้นตั้งใจ
จะวัด ความสัมพันธ์นั้นแทนได้ด้วยฟังก์ชันการตอบข้อคำถาม (item response function)
หรือ เส้นแนวทาง (trace line) หรือในกรณีที่มีคุณลักษณะ (latent trait)
เดียว ฟังก์ชันนี้เรียกว่าโค้งแสดงลักษณะข้อคำถาม (item characteristic curve--icc)
เมื่อมีฟังก์ชันหรือโค้งนี้ไว้แล้ว เราจะสามารถใช้โค้งนี้ทำนายการตอบข้อคำถามข้อใด ๆ ของ
ผู้สอบคนหนึ่งได้เมื่อทราบระดับของคุณลักษณะของผู้สอบคนนั้น

ถ้ากล่าวในเชิงคณิตศาสตร์ กำหนดให้ θ แทนคุณลักษณะที่ต้องการวัด
คุณลักษณะนี้อาจเป็นความสามารถ ทักษะ ทักษะ ทักษะ หรือลักษณะทางจิตใจ ๆ แต่ในงานนี้ศึกษา
แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ฉะนั้น จึงหมายถึงความสามารถใน
วิชาคณิตศาสตร์หรือผลสัมฤทธิ์วิชาคณิตศาสตร์ ข้อคำถามในแบบทดสอบนี้เป็นแบบที่ตรวจให้
คะแนนเป็นสองทาง นั่นคือตอบถูกให้คะแนน 1 ตอบผิดให้คะแนน 0 ใช้สัญลักษณ์ u_i
แทนคะแนนของข้อ i และ $\mathbf{u}' = (u_1, \dots, u_i, \dots, u_n)$ แทนชุดของคะแนนหรือแบบแผน
การตอบจากแบบทดสอบที่มี n ข้อ ในกรณีที่ข้อคำถามเป็นแบบที่ตรวจให้คะแนนเป็นสอง
ทางนี้ฟังก์ชันการตอบข้อคำถามมีชื่อเรียกโดยเฉพาะว่าโค้งแสดงลักษณะข้อคำถาม (icc)
คะแนนรายข้อ u_i เกี่ยวข้องกับความสามารถ θ ตามฟังก์ชันที่ให้ความน่าจะเป็นของ

คะแนนจากข้อหนึ่งของผู้ตอบคนหนึ่ง (ที่เลือกมาอย่างสุ่ม) ที่มีความสามารถระดับหนึ่ง
ที่กำหนดให้ ฟังก์ชันดังกล่าวคือ

$$Q_i(\theta) = \text{Prob}(U_i=0|\theta)$$

U_i และ 0 อยู่ในรูปแมทริกซ์

$$P_i(\theta) = \text{Prob}(U_i=1|\theta) = 1 - Q_i(\theta)$$

และเราสามารถสร้างฟังก์ชันการแจกแจงความน่าจะเป็นของ U_i ได้ดังนี้

$$f_i(u_i|\theta) = \text{Prob}(U_i=u_i|\theta) = P_i(\theta)^{u_i} Q_i(\theta)^{1-u_i} = \begin{cases} P_i(\theta) & \text{ถ้า } u_i=1 \\ Q_i(\theta) & \text{ถ้า } u_i=0 \end{cases}$$

นอกจากนี้ฟังก์ชันถดถอยของผลการตอบข้อใด ๆ u_i ก็คือ icc ของข้อ
นั้นนั่นเอง เนื่องจาก

$$\Sigma(U_i|\theta) = 1 \cdot f_i(1|\theta) + 0 \cdot f_i(0|\theta) = P_i(\theta)$$

ส่วนใหญ่เราสนใจกรณีที่ $P_i(\theta)$ เพิ่มขึ้นตาม θ ซึ่งในกรณีนี้ u_i เป็นตัว
ชี้และตัววัด θ

ฟังก์ชันการตอบข้อคำถามเป็นความน่าจะเป็น P หรือ $P(\theta)$ ของผู้สอบที่มี
ความสามารถระดับ θ_k ในการตอบข้อนั้นได้ถูก โดยปกติถ้าข้อคำถามมีอำนาจจำแนกอยู่บ้าง
ความน่าจะเป็น $P(\theta)$ เพิ่มขึ้นเมื่อ θ สูงขึ้น เบิร์นบอม (Birnbau, 1968) เสนอ
ฟังก์ชันโลจิสติก (logistic function) เพื่อใช้บรรยาย $P(\theta)$ ฟังก์ชันทางคณิตศาสตร์
นี้สามารถเขียนแทนในรูปดังนี้ (Birnbau, 1968; p.400)

$$P_i(\theta_k) = \frac{1}{1 + e^{-1.7a_i(\theta_k - b_i)}} \quad (1)$$

เมื่อ i แทนข้อคำถามข้อที่ i และ $i = 1, 2, 3, \dots, n$ a_i และ b_i เป็น

พารามิเตอร์ของข้อ ซึ่งเป็นอำนาจจำแนกและเป็นความยากของข้อตามลำดับ 1.7

เป็นค่าที่ช่วยให้โมเดลปกติสะสม และโมเดลโลจิสติกใกล้เคียงกันที่สุด e เป็นค่าคงที่

2.71828...

$P_{ik} = P_i(\theta_k)$ เป็นความน่าจะเป็นที่ผู้สอบคนที่ k จะตอบข้อ i ได้ถูก

หรือ เป็นความน่าจะเป็นที่ข้อ i จะถูกตอบได้ถูกต้องโดยผู้สอบคนหนึ่งที่มีเลือกมา

โดยที่ผู้สอบคนนี้มีความสามารถ $\theta = \theta_k$ การตีความดังที่กล่าวมาแล้วนี้ ไม่ได้กล่าว

ถึงความน่าจะเป็นที่ผู้สอบคนหนึ่งที่กำหนด จะตอบข้อคำถามข้อที่กำหนดได้ถูกต้อง

icc ในรูปของโลจิสติก ประมาณรูปโค้งปกติได้ใกล้เคียงแต่

ง่ากว่าในลักษณะที่ว่าข้อมูลทั้งหมดเกี่ยวกับ θ ที่ได้จากแบบแผนการตอบ (v) อยู่

ในสูตรดังนี้ (Birnbbaum, 1968, p.401)

$$x = x(v) = \sum_{i=1}^n a_i u_i$$

ซึ่งไม่ขึ้นอยู่กับการพารามิเตอร์ความยาก b_i

ฟังก์ชันที่แสดงในสมการ (1) รู้จักกันทั่วไปภายใต้ชื่อโมเดลโลจิสติกที่มี

พารามิเตอร์สองตัว (two-parameter logistic model--2P logistic model)

ถ้าอำนาจจำแนกของทุกข้อมีค่าเท่ากัน สมมติว่า $a_i = 1$ และให้ $D=1$

เราจะสามารถเขียนฟังก์ชันในสมการ (1) ให้อยู่ในรูปที่ง่ายขึ้นดังนี้ (Birnbbaum,

1968 p.402)

$$P_i(\theta_k) = \frac{1}{1 + e^{-(\theta_k - b_i)}} = \frac{1}{1 + e^{(b_i - \theta_k)}} \quad (2)$$

โมเดลนี้เป็นกรณีพิเศษของโมเดลในสมการ (1) ซึ่งทุกข้อมีอำนาจจำแนกเท่ากัน มีเพียงความยากเท่านั้นที่แปรเปลี่ยนนี้คือ โมเดลโลจิสติกที่มีพารามิเตอร์ตัวเดียว (one-parameter logistic model-- โมเดล 1PL)) ราช (Rasch, 1960) เป็นผู้พัฒนาโมเดลนี้ขึ้นมาโดย อาศัยเหตุผลที่แตกต่างจากเบิร์นบอม

ในบางครั้งผู้สอบที่มีความสามารถต่ำมาก ๆ จะตอบข้อคำถามแบบเลือกตอบได้โดยการเดาจึงมีการเสนอโมเดลสำหรับข้อคำถามเช่นนี้ โมเดลนี้สมมุติว่า ถ้าผู้สอบคนหนึ่งมีความสามารถระดับ θ เขาจะรู้คำตอบที่ถูกด้วยความน่าจะเป็นที่ได้จากฟังก์ชันปกติสะสม (normal ogive function) $\phi[a_i(\theta-b_i)]$ นอกจากนี้ยังสมมุติอีกว่า ถ้าเขาไม่รู้คำตอบถูกเขาจะเดา และความน่าจะเป็นที่เขาจะเดาถูกคือ c_i ดังนั้นความน่าจะเป็นในการตอบถูกข้อใดข้อหนึ่งคือ icc นี้ (Birnbbaum, 1968, p.404).

$$P_i(\theta) = c_i + (1-c_i) \phi\{a_i(\theta-b_i)\} \quad (3)$$

เมื่อ $\phi(x) = \int_{-\infty}^x \varphi(t)dt = \int_{-\infty}^x \varphi(t) dt$ และ $\varphi(t)$ เป็นฟังก์ชันความถี่ปกติ (normal frequency function)

ฟังก์ชันในสมการ (3) มีค่าเข้าสู่ค่า c_i เมื่อ θ ลดลง

ในทำนองเดียวกันโมเดลโลจิสติกอาจพิจารณาความน่าจะเป็นในการเดาได้โดยปรับ icc ดังนี้ (Lord, 1980, p.12; พงษ์จิต อินทสุวรรณ 2528, หน้า 15)

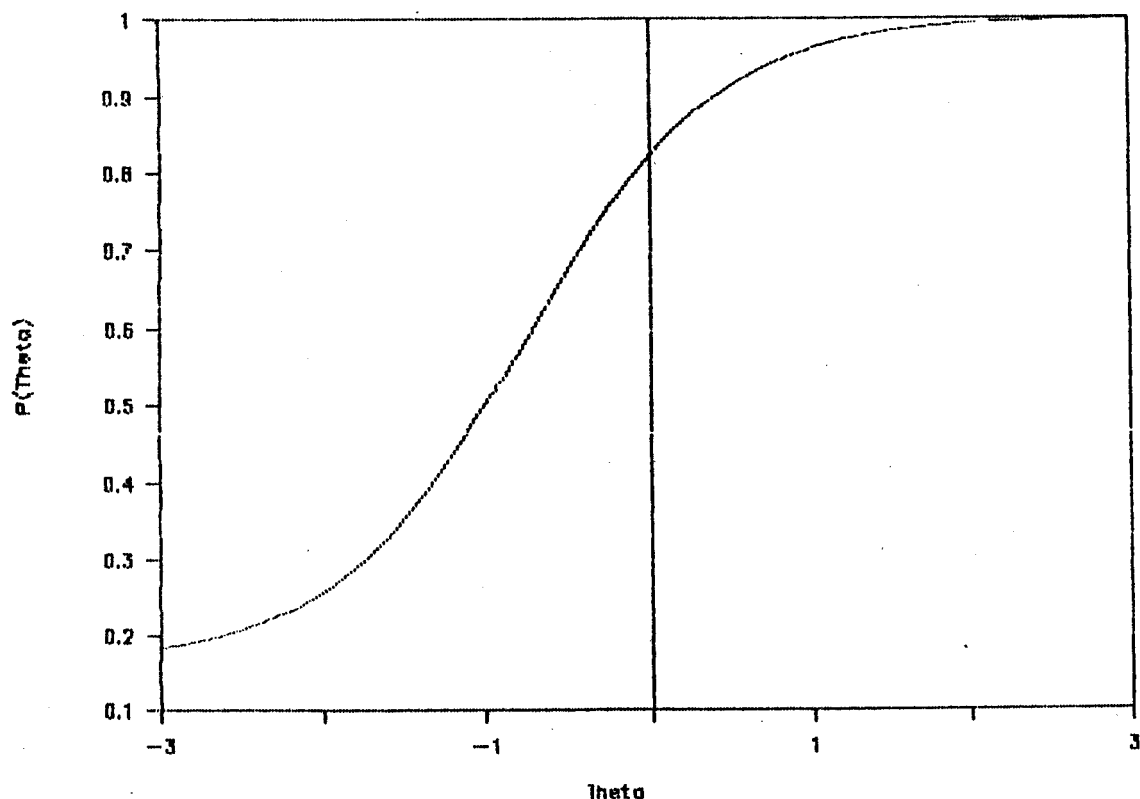
$$P_i(\theta) = c_i + \frac{1+c_i}{1+e^{-1.7a_i(\theta-b_i)}} \quad (4)$$

โมเดลนี้มีพารามิเตอร์ 3 ตัวจึงเรียกว่าโมเดล 3PL และ icc สำหรับโมเดลนี้สามารถเขียนแสดงรูปได้ดังนี้

จากภาพ 1 พารามิเตอร์ c เป็นความน่าจะเป็นที่บุคคลหนึ่งซึ่งมีความสามารถต่ำมากจะตอบข้อนั้นได้ถูก นั่นคือ icc เข้าสู่ค่าต่ำสุด c_1 เมื่อ θ ลดลง พารามิเตอร์ตัวนี้ถูกเรียกว่า พารามิเตอร์ของการเดา (guessing parameter) (Lord, 1980) ช่วงของแกนตั้งเปลี่ยนจาก 0 ถึง 1.0 เป็นช่วง c ถึง 1.0 ถ้าข้อคำถามมีตัวเลือก 4 ตัวเลือก และผู้ตอบเลือกตอบ 1 ตัวเลือกอย่างสุ่ม ซึ่งเป็นการเดา ค่า $c_1 = 1/4$ ข้อใดไม่สามารถตอบถูกต้องด้วยการเดา $c_1 = 0$ พารามิเตอร์ a_1 เป็นสัดส่วนกับความชันของโค้งที่จุดเปลี่ยนโค้ง ความชันนี้มีค่า $.425 a_1 (1 - c_1)$ พารามิเตอร์ a_1 แทนอำนาจจำแนก (discriminating power) ของข้อ พารามิเตอร์ตัวนี้เป็นตัวบอกว่า icc แปรไปตามระดับของ θ มากน้อยเพียงใด ถ้า a_1 มีค่าสูงแสดงว่า icc เปลี่ยนแปลงเร็วเมื่อค่า θ เปลี่ยนไปและโค้งมีลักษณะชัน ในทางตรงกันข้ามถ้า a_1 มีค่าต่ำแสดงว่า icc เปลี่ยนแปลงช้าเมื่อค่า θ เปลี่ยนไป และโค้งมีลักษณะลาด

พารามิเตอร์ b_1 แทนความยากของข้อ (item difficulty) เป็นตัวกำหนดตำแหน่งของโค้งบนมาตรา θ ข้อคำถามที่ยาก ค่า b_1 มีค่าสูงและโค้งจะมีตำแหน่งค่อนข้างสูงของแกน θ ส่วนข้อที่ง่าย ค่า b_1 มีค่าต่ำ และโค้งจะเลื่อนไปทางซ้าย จุดเปลี่ยนโค้งของฟังก์ชันโลจิสติกอยู่ที่ $\theta = b$ เมื่อไม่มีการเดา ($c_1 = 0$) ค่า b_1 จะตรงกับระดับความสามารถของบุคคลที่ตอบข้อนั้นถูกต้องด้วยความน่าจะเป็น .5 แต่ถ้ามีการเดา ($c_1 > 0$) ค่า b_1 จะตรงกับระดับความสามารถของบุคคลที่ตอบข้อนั้นถูกต้องด้วยความน่าจะเป็นที่มีค่าอยู่ระหว่าง c_1 และ 1.0

ลอร์ด (Lord, 1980) ได้แสดงการตรวจสอบว่าโมเดลในสมการ (4) เพียงพอที่จะบรรยายสถานการณ์ที่เป็นจริง โดยเปรียบเทียบโค้งโลจิสติกที่มีรูปฟังก์ชันในสมการ (4) กับโค้งการถดถอยของคะแนนรายข้อลงบนความสามารถ สำหรับข้อคำถามชุดเดียวกัน ซึ่งวัดความถนัดด้านภาษา ปรากฏว่าโค้งทั้งสองมีความใกล้เคียงกัน



ภาพ 1 โค้งแสดงลักษณะของข้อคำถาม

นอกจากนั้นยังเสนอหลักฐานที่แสดงว่าสัดส่วนการตอบถูกเป็นฟังก์ชันที่เพิ่มขึ้นตามคะแนนรวม แม้ว่าผู้สอบบางคนจะมีโอกาสเดาถูกได้สูงกว่า $1/A$ เมื่อ A แทนจำนวนตัวเลือกสำหรับแต่ละข้อ

ข้อตกลงเบื้องต้นที่สำคัญ

การพัฒนาโมเดลการตอบข้อคำถามอาศัยข้อตกลงเบื้องต้นสองประการคือ สมมติว่าข้อคำถามทุกข้อในแบบทดสอบฉบับเดียวกันเป็นเอกพันธ์ (homogeneous items) ในลักษณะที่ว่าทุกข้อวัดความสามารถตัวเดียวกันเพียงตัวเดียวเท่านั้น จากฟังก์ชันการตอบข้อคำถามในสมการ (1)-(4) แสดงว่าความน่าจะเป็นของการตอบข้อคำถามข้อหนึ่งได้ถูกขึ้นอยู่กับพารามิเตอร์ของข้อ (อาจมี 1 หรือ 2 หรือ 3 ตัว) กับระดับความสามารถของผู้ตอบเท่านั้น ไม่มีสิ่งอื่นใดอีกเลย ดังนั้นภายใต้โมเดลดังกล่าว เมื่อกำหนดข้อคำถาม

ข้อหนึ่งให้แล้ว เราเพียงแต่ทราบระดับความสามารถ θ ของผู้สอบเท่านั้นเราจะสามารถบอกได้ว่าเขาจะมีความน่าจะเป็นเท่าไรในการที่จะตอบข้อนั้นได้ถูก

ตัวอย่างของแบบทดสอบที่น่าจะเป็นเอกพันธ์ (มีมิติเดียว) คือ แบบทดสอบคำศัพท์ อนุกรม ตัวเลข การสะกดคำ มิติสัมพันธ์ เป็นต้น ส่วนแบบทดสอบคณิตศาสตร์ที่ประกอบด้วยข้อคำถามที่เป็นเหตุผลทางคณิตศาสตร์เสียครั้งหนึ่ง และ เป็นพีชคณิตอีกครั้งหนึ่ง ย่อมจะมีอย่างน้อยสองมิติ ความเป็นมิติเดียวนั้นเป็นจริงได้ยากสำหรับแบบทดสอบที่มีจริงเป็นเพียงการประมาณเพื่อผลในทางปฏิบัติ (Lord, 1968)

คุณสมบัติอย่างหนึ่งของแบบทดสอบที่มักจะเข้าใจผิดกันว่ามีความหมายเหมือนลักษณะเอกพันธ์ของแบบทดสอบคือ ความคงเส้นคงวาภายใน (internal consistency) แบบทดสอบที่มีความคงเส้นคงวาภายใน ประกอบด้วยข้อคำถามต่าง ๆ ที่มีความสัมพันธ์กันสูง แต่ทุกข้อไม่จำเป็นต้องวัดมิติเดียวกัน ส่วนแบบทดสอบที่มีความเป็นเอกพันธ์นั้น หมายถึงทุกข้อในแบบทดสอบวัดมิติเดียวกัน ดังนั้นแบบทดสอบที่มีความเป็นเอกพันธ์จึงมีความคงเส้นคงวาภายในสูงด้วย แต่แบบทดสอบที่มีความคงเส้นคงวาภายในสูงไม่จำเป็นต้องมีความเป็นเอกพันธ์ (Green, et al., 1977, p.831)

สัมประสิทธิ์แอลฟา (coefficient alpha) เป็นค่าประมาณ (ขอบเขตล่าง) ความคงเส้นคงวาภายในของแบบทดสอบ (Lord, 1980, p. 8; พงจิต อินทสุวรรณ, 2528, หน้า 9) ค่าสัมประสิทธิ์แอลฟาสูงจึงไม่จำเป็นต้องหมายถึงว่าแบบทดสอบนั้นมีความเป็นเอกพันธ์หรือมีมิติเดียว (unidimension)

แม็กโดนัล (McDonald, 1981, p.113) สรุปว่าแบบทดสอบที่ประกอบด้วยข้อคำถามที่ตรวจให้คะแนนเป็น 2 แบบ มีมิติเดียวก็ต่อเมื่อข้อคำถามเหล่านั้นเหมาะกับโมเดลที่มีแฟคเตอร์ร่วม (common factor model) (โดยทั่วไปไม่ใช่โมเดลเชิง

เส้นตรง) ที่มีแพคเตอร์ร่วมกันเดี๋ยวนั้น

ข้อตกลงเบื้องต้นประการที่สองคือ ความเป็นอิสระเฉพาะที่ (local independence) ข้อตกลงเบื้องต้นนี้กล่าวว่า ความน่าจะเป็นในการตอบข้อ i ได้ถูก เมื่อกำหนด θ ให้ เท่ากับความน่าจะเป็นในการตอบข้อ i ได้ถูก เมื่อกำหนด θ และ ผลการตอบข้อ $j, k, \dots$ ของผู้สอบให้ด้วย นั่นหมายความว่าความสามารถ θ เท่านั้นที่สำคัญในการจะบอกว่าเรามีโอกาสเท่าไรที่จะประสบความสำเร็จในการตอบข้อ i การทราบว่าเราประสบความสำเร็จหรือไม่ในการตอบข้อ $j, k, \dots$ ไม่ได้เพิ่มประโยชน์อันใด (ในความหมายนี้ข้อตกลงเบื้องต้นเกี่ยวกับความเป็นอิสระเฉพาะที่ จึงตรงกับข้อตกลงเบื้องต้นของการมีมิติเดียว จึงกล่าวว่าข้อตกลงเรื่องความเป็นอิสระเฉพาะที่ได้มาจากลักษณะความเป็นมิติเดียวโดยลอร์ดโนมิตี (Lord, 1980, p. 19; ผจจจิต อินทสุวรณ์ 2528 หน้า 25) เมื่อเป็นดังนี้สำหรับผู้สอบคนหนึ่ง คำตอบข้อหนึ่งของเขาส่อมเป็นอิสระเต็มที่จากคำตอบของเขาเมื่อตอบข้ออื่นใดในแบบทดสอบฉบับเดี๋ยวนั้น หรือกล่าวได้ว่าในกลุ่มข้อใด ๆ ของประชากรผู้สอบ ซึ่งทุกคนมีความสามารถระดับเดี๋ยวกัน คะแนนที่ได้จากข้อหนึ่งจะเป็นอิสระโดยสิ้นเชิงจากคะแนนในข้ออื่นใด แต่สำหรับในกลุ่มผู้สอบทั่ว ๆ ไปซึ่งมักจะประกอบด้วยบุคคลที่มีความสามารถแตกต่างกันไป คะแนนจากข้อต่าง ๆ ส่อมสัมพันธ์กัน

ถ้านิยามคุณลักษณะ (trait) หรือความสามารถได้แคบและชัดเจนจนกระทั่งสามารถสร้างข้อคำถามต่าง ๆ ขึ้นมาวัดคุณลักษณะนั้นได้ และคุณลักษณะหรือความสามารถนั้นคงที่ (stable) อยู่ระยะหนึ่งสำหรับประชากรกลุ่มใหญ่ แล้วพารามิเตอร์ของข้อไม่ขึ้นอยู่กับกลุ่มบุคคลที่สอบ นั่นคือพารามิเตอร์ของข้อหนึ่ง ๆ ไม่แปรปรวน (invariance) เมื่อประมาณจากกลุ่มตัวอย่างที่มีความสามารถต่างกัน ภายในขอบเขตของประชากร

นอกจากนี้พารามิเตอร์ความสามารถของผู้สอบไม่ขึ้นอยู่กับชุดของข้อคำถามที่ใช้สอบ ซึ่งเลือกมาจากประชากรของข้อคำถามที่วัดมิติเดียวกัน (Lord and Novick, 1968, p.379; Lord, 1980, และ ผจจจิต อินทสุวรรณ 2528. ตอน 3.4) คุณสมบัตินี้เป็นประโยชน์อย่างมาก ในการวัดผลปัจจุบัน เนื่องจากเราสามารถประมาณค่าพารามิเตอร์ของข้อคำถามจากคนกลุ่มใดในประชากรก็ได้ อย่างไรก็ดีเบคเกอร์ (Baker, 1977, p.170-172) ตั้งข้อสังเกตว่า คุณสมบัติดังกล่าวนี้เป็นไปได้ในทางปฏิบัติ เมื่อใช้กลุ่มตัวอย่างของผู้สอบขนาดใหญ่ เพื่อประมาณค่าพารามิเตอร์ของข้อคำถามชุดหนึ่ง (pool) แล้วให้ค่าที่ประมาณได้เป็นพารามิเตอร์ของข้อ ต่อไปจะ
ใช้ข้อคำถามชุดย่อยชุดใดก็ได้จากชุดใหญ่เดิม เพื่อให้ประมาณค่าความสามารถของผู้สอบ
วิธีการในการตรวจสอบความเป็นมิติเดียวของข้อคำถาม

วิธีการตรวจสอบความเป็นมิติเดียวของข้อคำถามต่าง ๆ ในแบบทดสอบฉบับหนึ่ง ลอร์ด (Lord, 1980 p.21) เสนอแนะวิธีคร่าว ๆ ไว้ดังนี้ คำนวณค่าเลเท็นรุต (latent roots) หรือค่าไอเกน (eigen values) ของเมทริกซ์ของสหสัมพันธ์ เทตระคอรีระหว่างข้อที่มีค่าประมาณของคอมมิวนอลิตี้ (communalities) อยู่บนแนวเส้นทะแยงมุม ถ้าผลที่ได้พบว่าเลเท็นรุตตัวแรกมีค่าสูง เมื่อเปรียบเทียบกับค่าเลเท็นรุตตัวที่สอง และค่าเลเท็นรุตตัวที่สองนี้มีค่าไม่แตกต่างจากค่าเลเท็นรุตตัวอื่น ๆ มากนัก ผลเช่นนั้นแสดงว่ามีองค์ประกอบร่วม (common factor) องค์ประกอบเดียวที่สำคัญ เราจะสามารถสรุปได้ว่า ข้อคำถามเหล่านั้นมีมิติเดียวโดยประมาณ ในการศึกษาครั้งนี้ จะใช้วิธีการนี้ เบิร์นบอม (Birnbaum, 1968, p.381) ยังกล่าวถึงอีกวิธีการหนึ่งที่นักวิชาการจำนวนหนึ่งใช้กันคือ วิเคราะห์องค์ประกอบจากสหสัมพันธ์ เทตระคอรีระหว่างข้อของแบบทดสอบฉบับต่าง ๆ เพื่อที่จะดูว่ามีองค์ประกอบร่วมมากกว่าหนึ่งองค์ประกอบ

หรือไม่ ซึ่งเมื่อพิจารณาส่วนที่เหลือ (residuals) หลังจากแยกหนึ่งแฟคเตอร์ออกไปแล้ว ปรากฏว่าค่าเหล่านี้มีค่าน้อยจนถือว่าเป็นความบังเอิญได้ วิธีการพิจารณาส่วนที่เหลือนี้ตรงกับข้อเสนอแนะของ คีทส์ (Keats, 1989) แต่คีทส์ใช้สหสัมพันธ์ฟี (phi correlations) แทนสหสัมพันธ์เทตระคอร์ริค

ความเป็นมิติเดียวของแบบทดสอบมีความหมายกว้างกว่าองค์ประกอบร่วมขององค์ประกอบเดียว (Birnbaum, 1968; Lord, 1980, p.19; และ ผจวจิต อินทสุวรณ์, 2528, หน้า 25) ด้วยเหตุผลที่ว่า การวิเคราะห์องค์ประกอบต้องมีเงื่อนไขต่าง ๆ ซึ่งทฤษฎีการตอบข้อคำถามไม่ต้องมี กล่าวคือ ถ้าความสามารถ θ มีการแจกแจงเป็นปกติ ฟังก์ชันการตอบข้อคำถามเป็นแบบปกติสะสม และไม่มีการเดาแล้ว สหสัมพันธ์เทตระคอร์ริคระหว่างข้อจะมีแฟคเตอร์ร่วมเพียงแฟคเตอร์เดียว ถ้าเงื่อนไขต่าง ๆ ไม่เป็นดังที่กล่าว การวิเคราะห์องค์ประกอบจะไม่ให้แฟคเตอร์ร่วมเพียงแฟคเตอร์เดียว แต่แบบทดสอบยังคงมีมิติเดียวได้

นอกจากนี้ การวิเคราะห์องค์ประกอบในเชิงเส้นตรงไม่เหมาะกับการนำมาใช้ตรวจสอบความเป็นมิติเดียวของข้อมูลที่จะใช้โมเดลโลจิสติก ในโมเดลโลจิสติกไม่จำเป็นที่ความสัมพันธ์ระหว่างคะแนนจากแบบทดสอบ กับคุณลักษณะ (trait) ที่วัดจะต้องเป็นเชิงเส้นตรง ดังนั้นจึงมีผู้สนใจศึกษาวิธีการต่าง ๆ ในการตรวจสอบความเป็นมิติเดียวของแบบทดสอบโดยใช้วิธีการอื่น ๆ รวมทั้งการวิเคราะห์องค์ประกอบที่ไม่ใช่โมเดลเชิงเส้นตรงด้วย

บีจาร์ (Bejar, 1980) เสนอวิธีการตรวจสอบมิติของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนโดยการประมาณค่าพารามิเตอร์ความยาก (b) ของแต่ละข้อสองครั้ง (sampling error ในการประมาณค่า b มีขนาดเล็กกว่าในการประมาณค่า a)

คือ ประมวลค่าโดยอาศัยคะแนนจากแบบทดสอบทั้งฉบับ และประมวลค่าโดยอาศัยคะแนนจากแต่ละเนื้อหา ถ้าข้อคำถามในแบบทดสอบมีมิติเดียว ค่าประมาณความยากทั้งสองครั้งควรเท่ากัน หรือเมื่อนำค่าความยากทั้งสองชุดมาพล็อตกราฟจุดที่แทนค่าประมาณความยากควรเข้าใกล้เส้นที่มีความชันเป็น 1 และตัดแกน y ที่ 0.0 หรือแสดงความสัมพันธ์เชิงเส้นตรง ถ้าค่าประมาณทั้งสองครั้งไม่เท่ากัน แสดงว่าผลการตอบข้อคำถามเหล่านั้นขึ้นอยู่กับข้ออื่น ๆ ที่ประกอบกันเป็นแบบทดสอบ ซึ่งไม่สอดคล้องกับข้อตกลงเบื้องต้นที่ว่ามีคุณลักษณะเดียวเท่านั้นที่อธิบายผลการตอบข้อคำถามทุกข้อ

แฮมเบิลตัน และ โรวินELLI (Hambleton and Rovinelli, 1986) สร้างข้อมูลขึ้น 5 ชุด ภาษาได้โมเดล 3PL โดยให้ข้อมูลมี 1 มิติหรือ 2 มิติ เขาใช้วิธีการตรวจสอบมิติของข้อมูล 4 วิธีด้วยกันคือ การวิเคราะห์องค์ประกอบในเชิงเส้นตรง (linear factor analysis) การวิเคราะห์องค์ประกอบที่ไม่เป็นเชิงเส้นตรง (nonlinear factor analysis--NLFA) การวิเคราะห์ส่วนที่เหลือ (residual analysis) และวิธีการของ เบจาร์ (Bejar's method) ผลปรากฏว่าวิธี NLFAสามารถระบุมิติของข้อมูลได้ถูกต้อง 3 ชุดจากทั้งหมด 5 ชุด ซึ่งนับว่าดีที่สุด ในจำนวนวิธีทั้งหมดที่ใช้

กล่าวอย่างย่อ ๆ หลักการวิเคราะห์ส่วนที่เหลือคือการให้ (fit) โมเดล IRT ที่มีมิติเดียวที่สนใจ กับข้อมูลจากผลการสอบ แล้ววิเคราะห์ความแตกต่าง (discrepancies) หรือส่วนที่เหลือ (residuals) ระหว่างค่าประมาณพารามิเตอร์จากโมเดล และข้อมูลของข้อคำถาม วิธีการคือแบ่งคะแนน θ ออกเป็นช่วงเท่า ๆ กัน ผู้สอบแต่ละคนจะถูกจัดลงในช่วงต่าง ๆ ตามค่าประมาณ θ ของเขาในแต่ละข้อ และแต่ละกลุ่มผู้สอบที่ตกลงในแต่ละช่วงของ θ มีการเปรียบเทียบระหว่างผลการสอบจริง

(สัดส่วนการตอบถูก) กับสัดส่วนการตอบถูกที่ประมาณจาก icc ที่ใช้ ความแตกต่างนี้ เรียกว่าส่วนที่เหลือ ซึ่งถือเป็นส่วนที่เหลือดิบ และเมื่อหารด้วยความคลาดเคลื่อนมาตรฐานของค่าประมาณสัดส่วนการตอบถูกก็กลายเป็นส่วนที่เหลือมาตรฐาน (standardized residual--SR) ถ้าโมเดลที่มีมิติเดียวที่เลือกมาใช้เหมาะกับข้อมูล ค่า SR เหล่านี้จะมีค่าต่ำและกระจายอย่างสม่ำเสมอใกล้ ๆ ค่า 0.0

ผลการศึกษาที่แสดงให้เห็นว่าการวิเคราะห์องค์ประกอบที่ไม่ใช่เชิงเส้นตรงสามารถใช้ตรวจสอบมิติของข้อคำถามได้ดีซึ่งปรากฏในงานของ คิงสตัน (Kingston, 1986) ซึ่งใช้โปรแกรมคอมพิวเตอร์ TESTFACT ข้อมูลที่นำมาประเมินมิติเป็นการตอบ GMAT (Graduate Management Admission Test) ฉบับภาษา (Verbal) และ ตัวเลข (Quantitative) โปรแกรมนี้ทำ full information factor analysis วิธีการใน TESTFACT คือทำ stepwise factor analysis นั่นคือ ขั้นแรกทำ แฟกเตอร์เดียว (one-factor solution) ต่อไปทำสองแฟกเตอร์ (two-factor solution) แล้วคำนวณความแตกต่างระหว่างไค-สแควร์ (chi-squares) สำหรับ ผลการทำแฟกเตอร์ทั้งสองครั้ง เป็นการทดสอบว่าจำนวนแฟกเตอร์ที่เพิ่มขั้นนั้น มีนัยสำคัญทางสถิติหรือไม่ อาจเพิ่มแฟกเตอร์ที่ 3, 4, ฯลฯ ได้อีกถ้าจำเป็น ผลการวิเคราะห์ ปรากฏว่า สำหรับฉบับภาษานั้นแฟกเตอร์ที่ 1, 2, และ 3, สามารถอธิบายความแปรปรวนของคะแนนได้ร้อยละ 21.3 3.9 และ 3.2 ตามลำดับ ส่วนฉบับตัวเลขมี ร้อยละของความแปรปรวนที่แฟกเตอร์ที่ 1, 2, และ 3 อธิบายได้เป็น 33.7, 3.5, และ 1.3 ตามลำดับ ผลเหล่านี้แสดงว่าแฟกเตอร์แรกเด่นที่สุด และสามารถสรุปได้ว่า GMAT มีมิติเดียว ผลสรุปนี้สอดคล้องกับการวิเคราะห์องค์ประกอบเชิงเส้นตรงแต่ชัดเจนกว่า

มีหลักฐานที่แสดงแล้วว่า โปรแกรมคอมพิวเตอร์ LOGIST ใช้ประมาณค่าพารามิเตอร์ได้ดีแม้เมื่อข้อมูลไม่ได้มีองค์ประกอบเดียวโดยแท้ (Drasgow and Parsons, 1983)

การประมาณค่าพารามิเตอร์

สำหรับโมเดลโลจิสติกที่มีพารามิเตอร์สามตัว (โมเดล 3PL) ความน่าจะเป็นที่ผู้สอบคนที่ k จะตอบข้อที่ i ได้ถูกต้องคือ สมการ (4) หรือ

$$P_i(\theta_k) = c_i + \frac{1-c_i}{1+\exp\{-1.7a_i(\theta_k - b_i)\}} \quad (5)$$

เมื่อ θ_k เป็นค่าความสามารถของผู้สอบคนที่ k และ a_i , b_i , และ c_i เป็นอำนาจจำแนก ความยาก และแอซิมโทตล่างของข้อที่ i

กำหนด θ_k (ความสามารถของผู้สอบ k) ให้ การแจกแจงที่มีเงื่อนไขของคำตอบต่อข้อ i คือ

$$L_k(u_{ik} | \theta_k) = \frac{P_{ik}^{u_{ik}} Q_{ik}^{1-u_{ik}}}{Q_{ik}} \quad (6)$$

เมื่อ $Q_{ik} = 1 - P_{ik}$

u_{ik} เป็นผลการตอบข้อ i ของผู้สอบ k สำหรับคำตอบถูก $u_{ik} = 1$

ส่วนคำตอบผิด $u_{ik} = 0$

เมื่อสมมติว่ามีความเป็นอิสระเฉพาะที่ ความสำเร็จในการตอบข้อหนึ่งย่อม เป็นอิสระจากความสำเร็จในการตอบข้ออื่น ๆ เมื่อเขาตอบทุกข้อไม่มีเว้น ดังนั้นการแจกแจงร่วมของผลการตอบทุกข้อสำหรับผู้สอบคนหนึ่ง (คนที่ k) จึงเป็น (Lord, 1980, p.54; พงจิต อินทสุวรรณ 2528, หน้า 80)

$$L(u | \theta_k, a, b, c) = \prod_{i=1}^n \frac{P_{ik}^{u_{ik}} Q_{ik}^{1-u_{ik}}}{Q_{ik}} \quad (7)$$

เมื่อ $\tilde{u}$ เป็นเวกเตอร์ชนิดคอลัมน์ $\{u_{1k}, u_{2k}, \dots, u_{nk}\}$ และ $\tilde{a}, \tilde{b}, \tilde{c}$ เป็นเวกเตอร์ของพารามิเตอร์ของข้อที่ i

ถ้าทราบเวกเตอร์ของผลการตอบ ($\tilde{u}_i$) ของผู้ตอบคนที่ k สมการ (7) เป็นฟังก์ชันไลค์ลิฮูด (likelihood function) สำหรับความสามารถ θ_k เมื่อทราบค่าพารามิเตอร์ของข้อคำถามแล้ว และ θ_k ซึ่งเป็นค่าประมาณไลค์ลิฮูดสูงสุดของ θ_k ให้ค่าสูงสุดของฟังก์ชันไลค์ลิฮูดนี้

ถ้าสุ่มผู้สอบมา N คน จากประชากรกลุ่มใหญ่กลุ่มหนึ่ง แล้วการแจกแจงของ $\tilde{u}$ สำหรับผู้สอบแต่ละคนจะเป็นอิสระต่อกัน ดังนั้นการแจกแจงร่วมของผลการตอบทุกข้อ (n ข้อ) สำหรับผู้สอบทุกคนจึงเป็น (Lord, 1980, p.56; พงษ์จิต อินทสุวรรณ, 2528, หน้า 82)

$$L(U|\theta; \tilde{a}, \tilde{b}, \tilde{c}) = \prod_{k=1}^N \prod_{i=1}^n P_{ik}^{u_{ik}} Q_{ik}^{1-u_{ik}} \quad (8)$$

เมื่อ $P_{ik} \equiv P_i(\theta_k)$

U เป็นแมทริกซ์ของผลการตอบข้อคำถาม n ข้อสำหรับผู้ตอบ N คน

θ เป็นเวกเตอร์ $\{\theta_1, \theta_2, \dots, \theta_N\}$

สมการ (8) เป็นฟังก์ชันไลค์ลิฮูดของความสามารถ (θ) เนื่องจากทราบค่าพารามิเตอร์ของข้อแล้ว สมการนี้อาจเป็นฟังก์ชันไลค์ลิฮูดของทั้งพารามิเตอร์ของความสามารถ (θ) และพารามิเตอร์ของข้อ ($\tilde{a}, \tilde{b}, \tilde{c}$) ได้ เมื่อไม่ทราบพารามิเตอร์ของข้อมาก่อน ค่าประมาณไลค์ลิฮูดสูงสุด ($\tilde{\theta}$) และ $\tilde{a}, \tilde{b}, \tilde{c}$ ทั้งหมดรวมกันให้ค่าสูงสุดของฟังก์ชันในสมการ (8) ในกรณีนี้ต้องมีการประมาณค่าพารามิเตอร์พร้อม ๆ กันถึง $(N+3n)$ ตัวจากแมทริกซ์ขนาด $N \times n$ ของผลการตอบข้อคำถาม (U)

สำหรับผู้สอบหนึ่งคน เมื่อเลือกฟังก์ชันของ P_{ik} เป็นโลจิสติก ฟังก์ชันไลค์ลิฮูด (สมการ (7)) ย่อมเปลี่ยนไปเป็น

$$L(u|\theta_k) = \exp \left(\sum_{i=1}^n u_{ik} L_{ik} \right) \prod_{i=1}^n Q_{ik} \\ = \exp \left(-D \sum_{i=1}^n a_i b_i u_i \right) e^{D \theta_k S} \prod_{i=1}^n Q_{ik} \quad (9)$$

เมื่อ
$$S \equiv S(u) \equiv \sum_{i=1}^n a_i u_{ik} \quad (10)$$

มีหลักฐาน (Lord, 1980; พงจิต อินทสุวรรณ, 2528, หน้า 94-95) ที่แสดงให้เห็น

ว่า ถ้า (1) ฟังก์ชันการตอบข้อคำถามเป็นโลจิสติก (2) ไม่มีการเดา และ (3)

ทราบค่าพารามิเตอร์ของข้อแล้ว $S = \sum_{i=1}^n a_i u_{ik}$ เป็นสถิติที่เพียงพอในการประมาณค่าความสามารถ θ_k ของผู้สอบ

จากข้อความข้างบนย่อมแสดงให้เห็นว่า สำหรับโมเดล 3PL ค่าในสมการ

(5) ย่อมไม่มีสถิติที่เพียงพอสำหรับประมาณค่าพารามิเตอร์ความสามารถ ในกรณีที่ไม่มีสถิติที่เพียงพอวิธีการประมาณค่าพารามิเตอร์ที่นิยมใช้กันคือ วิธีไล่คัลลี้สูงสุด ค่าประมาณไล่คัลลี้สูงสุด $(\hat{\theta}_k)$ ($k = 1, 2, \dots, N$) และ $\hat{a}_i, \hat{b}_i, \hat{c}_i$ ($i = 1, 2, \dots, n$) ที่ประมาณจากแมทริกซ์ของผลการตอบข้อคำถาม (U) ค่าประมาณไล่คัลลี้สูงสุดได้มาจากการหาล็อกการวิธิมของฟังก์ชันไล่คัลลี้ของอนุพันธ์ของล็อกการวิธิมของฟังก์ชันนั้น แล้วกำหนดให้อนุพันธ์นั้นเท่ากับศูนย์ แล้วจึงแก้สมการไล่คัลลี้ชุดต่าง ๆ ที่ได้

ถ้าทราบค่าพารามิเตอร์ของข้อจากการทดลองสอบแล้ว ค่าประมาณความ

สามารถของผู้สอบแต่ละคนหาได้จากการแก้สมการดังกล่าว ซึ่งแต่ละสมการมีตัวแปรทราบ

ค่าเพียงตัวเดียว คือ θ_k ค่ารากของสมการดังกล่าวหาได้ด้วยวิธีหาค่าซ้ำ ๆ

(iterative procedures) แต่ค่ารากนี้อาจมีหลายค่าได้ถ้ามีจำนวนข้อน้อย (Samejima,

1973) อย่างไรก็ตามถ้าจำนวนข้อมีมากพอ สมการดังกล่าวจะให้ค่ารากเพียงค่าเดียว

ซึ่งค่ารากนี้จะเป็ค่าประมาณความสามารถที่คงเส้นคงวา

การประมาณค่าพารามิเตอร์โดยวิธีไลค์ลี่ยุติสูงสุดนี้มีกระบวนการที่ซับซ้อน ต้องใช้คอมพิวเตอร์ในการคำนวณ และสิ้นเปลืองค่าใช้จ่าย

วิธีที่ง่ายกว่า และสิ้นเปลืองน้อยกว่าวิธีการไลค์ลี่ยุติสูงสุดมีผู้เสนอไว้ดัง เช่น เจนเซมา (Jensema, 1976) เสนอวิธีประมาณค่าพารามิเตอร์ของข้อ (a_i และ b_i) โดยอาศัย (1) สหสัมพันธ์พ้อยท์ไบซีเรียล ระหว่างคะแนนรวมทั้งฉบับ (ที่ไม่รวมข้อ i) กับคะแนน (เป็นสองทาง) ผลการตอบข้อ i และ (2) จุดตัด (γ ดู Lord and Novick, 1968, p.370-371) ระหว่างการตอบถูก-ผิด สำหรับข้อ i บน response continuum ที่ไม่สามารถสังเกตได้ ทั้งนี้โดยถือว่าคะแนนรวมทั้งไม่รวมข้อ i แทนความสามารถ θ และถือว่ามีค่าประมาณที่ดีของ c อยู่แล้ว

บุคคลอีกคนหนึ่งคือ เยน (Yen, 1984) เสนอวิธีประมาณค่าพารามิเตอร์ ความสามารถของผู้สอบ โดยใช้วิธีไลค์ลี่ยุติสูงสุดจากจำนวน ค่าตอบถูก แทนที่จะใช้ แมตริกซ์ของค่าตอบรายข้อสำหรับผู้สอบแต่ละคน ซึ่งการใช้จำนวนค่าตอบถูกนั้นง่ายกว่า และไม่สิ้นเปลืองเท่ากับการใช้แมตริกซ์ดังกล่าว

วิธีทั้งสองให้ค่าประมาณพารามิเตอร์ที่ถูกต้องดีพอสมควรเมื่อตรวจสอบ โดยวิธีสร้างข้อมูลจำลองขึ้นมา และเทียบกับวิธีประมาณค่าโดยใช้ไลค์ลี่ยุติสูงสุด อย่างไรก็ตามวิธีทั้งสองนี้ไม่สามารถ ใช้แทนวิธีไลค์ลี่ยุติสูงสุดซึ่งอาศัยข้อมูลจาก แมตริกซ์ของผลการตอบรายข้อสำหรับผู้สอบแต่ละคนได้ ผลที่ได้จากวิธีทั้งสองนี้ กล่าวถึงนี้อาจใช้ช่วยในการคัดเลือกเบื้องต้น เพื่อการวิเคราะห์ต่อไปหรืออาจใช้เป็น ค่าเริ่มต้นในการประมาณค่าโดยวิธีไลค์ลี่ยุติสูงสุดได้

การประมาณค่าพารามิเตอร์เมื่อผู้สอบเว้นบางข้อ

ในการทำข้อสอบ ถ้าผู้สอบเว้นไม่ตอบ (ไม่รวมข้อที่ทำไม่ถึง) บางข้อ เราจะเว้นข้อเหล่านั้นในการประมาณค่าพารามิเตอร์โดยใช้สมการ (7) หรือ (8) เป็นการไม่ถูกต้อง เพราะเป็นการละเลยความรู้ที่ว่า การที่เขาเว้นไม่ตอบบางข้อ แสดงว่าเขาไม่รู้คำตอบที่ถูกต้องสำหรับข้อนั้น ๆ นั่นคือ โอกาสในการตอบถูกของเขามีเพียง c โดยประมาณ

ในกรณีของข้อสอบแบบเลือกตอบ (สำหรับในรายงานนี้เป็นเรื่องของข้อสอบแบบเลือกตอบ) ถ้าผู้สอบทำข้อสอบทุกข้อ การตรวจให้คะแนนโดยใช้จำนวนคำตอบถูก (x) เป็นสิ่งที่เหมาะสม แต่ถ้าผู้สอบเว้นบางข้อ การใช้คะแนนจำนวนคำตอบถูกจะกลายเป็นว่า ผู้สอบไม่ต้องการทำคะแนนให้ดีที่สุด สูตรการให้คะแนนนิยมใช้กันเพื่อทำโทษสำหรับการตอบผิด (Lord, 1974) คือ

$$y_k = x_k - w_k / (A-1) \quad (11)$$

เมื่อ A เป็นจำนวนตัวเลือกที่ให้ไว้ในข้อหนึ่ง ๆ

w_k เป็นจำนวนคำตอบผิดของผู้สอบ k

ผู้สอบจะเว้นข้อใดข้อหนึ่งถ้าเขาเชื่อว่าโอกาสที่จะตอบถูกไม่มากไปกว่า $c = 1/A$ ในทางตรงกันข้าม เนื่องจากข้อหนึ่งมี A ข้อเลือก โอกาสที่เขาจะตอบถูกไม่ควรจะน้อยกว่า c เนื่องจากเขาสามารถทำได้ด้วยการเดาสุ่ม ด้วยเหตุนี้จึงถือว่าถ้าผู้สอบคนหนึ่งจำเป็นต้องตอบข้อที่เขาเว้นไว้ทั้งหมด สัดส่วนในการตอบถูกก็จะเป็น c

การประมาณค่าพารามิเตอร์ในกรณีที่ผู้สอบเว้นไม่ทำบางข้อนั้น ลอร์ด (Lord, 1974, p.251) เสนอฟังก์ชันแทนโลจิสติกส์ยึดฟังก์ชันในสมการ (7) ดังนี้

$$L_k^* (y_{ik} | \theta_k) = \prod_{i=1}^n P_{ik}^v Q_{ik}^{1-v} \quad (12)$$

เมื่อ $\underline{v}_{1k}$ เป็นเวกเตอร์ของการตอบของผู้สอบ k

$$\underline{v}_{1k} = (v_{1k}, v_{2k}, \dots, v_{rk})'$$

$v_{1k} = 1, 0$, หรือ c ขึ้นอยู่กับคำตอบของเขาคือ, ผิด หรือ เว้น

ตามลำดับ ค่าต่าง ๆ ของ v_{1k} เป็นคะแนนรายชื่อที่ใช้ในการตรวจให้คะแนนตามสูตร

ใด ๆ คะแนน $\sum_{i=1}^n v_{ik}$ สัมพันธ์อย่างสมบูรณ์กับคะแนนจากสูตร (11)

ในกรณีที่ต้องประมาณค่าพารามิเตอร์ของข้อไปพร้อม ๆ กับความสามารถ

(θ_k) ต้องมีการวิเคราะห์ผลการตอบของผู้สอบจำนวนมาก (ควรเป็น 2 หรือ 3 พัน)

ในเวลาเดียวกัน สมการ (8) จึงถูกแทนด้วย

$$L^*(v|\theta) = \prod_{k=1}^N \prod_{i=1}^n \frac{v_{ik}}{p_{ik}} \frac{1-v_{ik}}{q_{ik}} \quad (13)$$

เมื่อ v เป็นเมทริกซ์ของผลการตอบที่สอบ n ข้อ สำหรับผู้สอบ N คน สมการนี้ใช้

หาค่าความสามารถ (θ_k) ของผู้สอบแต่ละคนในกลุ่มผู้สอบ N คน และค่าพารามิเตอร์

ทั้งสามของข้อแต่ละข้อ สำหรับข้อ n ข้อ เพื่อให้สมการ (13) มีค่าสูงสุด ค่าเหล่านี้

ถือเป็นค่าประมาณพารามิเตอร์ที่ต้องการ

ค่าที่ประมาณได้นี้ไม่ใช่ค่าประมาณไคลล์สูงสุด เนื่องจากสมการ (12)

และ (13) ไม่ใช่ฟังก์ชันไคลล์สูงสุด แทนค่าประมาณของ θ ด้วย θ^*

ลอร์ด (Lord, 1974) ทดลองใช้สมการ (13) เพื่อประมาณค่าพารามิเตอร์

ภายใต้โมเดล 3PL โดยอาศัยข้อมูลจริง 3 ชุด และเพื่อเป็นการเปรียบเทียบได้วิเคราะห์

ข้อมูลซ้ำโดยใช้สมการ (8) เมื่อแทนข้อที่เว้นไม่ตอบด้วยคำตอบชนิดสุ่ม ค่าประมาณชุดหลัง

นี้คือค่าประมาณไคลล์สูงสุด (MLE)

หลักฐานจากค่าพารามิเตอร์ของข้อทั้งสามตัว แสดงว่าวิธีการทั้งสองให้

ค่าที่ใกล้เคียงกันมากจนกล่าวได้ว่าเท่ากัน

นอกจากนี้ยังมีผลจากข้อมูลจำลองที่ชี้ให้เห็นว่าค่าประมาณจากทั้งสองวิธี

ใกล้เคียงกันมาก

วิธีการประมาณค่าพารามิเตอร์ในโปรแกรมคอมพิวเตอร์ LOGIST

โปรแกรมคอมพิวเตอร์ LOGIST (Wood, Wingersky, and Lord,

1976) เขียนด้วยภาษา FORTRAN IV ได้รับการพัฒนาขึ้นที่ศูนย์บริการด้านการทดสอบ
ทางการศึกษา (Educational Testing Service--ETS) ประเทศสหรัฐอเมริกา
เพื่อใช้ในการประมาณค่าพารามิเตอร์ของข้อ และความสามารของผู้สอบภายใต้
โมเดล 3PL วิธีการที่ใช้ในการประมาณค่าพารามิเตอร์ในโปรแกรมนี้อาศัยสมการ
(13) ผลการตอบข้อสอบนำมาลงโค้ดเป็น 4 แบบคือ ในกรณีที่ตอบถูก ตอบผิด เว้น
และทำไม่ถึง

สำหรับผู้สอบ N คน และจำนวนข้อ n ข้อ จำนวนพารามิเตอร์ที่จะต้อง
ประมาณค่ามีถึง $N + 3n$ ค่า การหาค่าประมาณกระทำโดยกระบวนการทำซ้ำ ๆ
(iterative procedure) โดยมีกระบวนการดังนี้ เริ่มแรก กำหนดพารามิเตอร์
ของข้อให้เป็นค่าคงที่ และประมาณค่าความสามารถสำหรับผู้สอบแต่ละคน เนื่องจาก
ความสามารถเหล่านั้นเป็นอิสระต่อกัน ต่อไปจึงหาค่าประมาณความสามารถนั้นเป็น
ค่าคงที่ และประมาณค่าพารามิเตอร์ของข้อสำหรับแต่ละข้อ เนื่องจากข้อคำถามต่าง ๆ
เป็นอิสระต่อกัน การกระทำทั้งสองตอนทีกล่าวแล้วรวมเรียกว่า 1 รอบ (stage)

กระบวนการประมาณค่าพารามิเตอร์แบ่งออกเป็น 4 ขั้น (steps)
แต่ละขั้นจบลงด้วยสภาพเข้าสู่จุดรวม (convergence) ก่อนที่จะเริ่มต้นต่อไป แต่ละขั้น
มีค่าที่เข้าสู่จุดรวมเมื่อฟังก์ชันที่ให้ (สมการ (13)) ให้ค่าประมาณพารามิเตอร์ใน 2 รอบ
แตกต่างกันน้อยกว่าเกณฑ์ที่กำหนดไว้

เพื่อให้ได้ผลดีที่สุด (ค่าเข้าสู่จุดรวม) ในการประมาณค่าพารามิเตอร์ จำนวนข้อคำถามไม่ควรต่ำกว่า 40 ข้อ และจำนวนผู้สอบไม่ควรต่ำกว่า 1,000 คน (Wood, et., 1976, p.5) นอกจากนี้โปรแกรมจะไม่ประมาณค่าความสามารถของผู้สอบที่ได้คะแนนเป็นศูนย์ ที่ได้คะแนนเต็ม และอาจไม่ประมาณค่าความสามารถของผู้ที่ตอบน้อยกว่า 1 ใน 3 ของจำนวนข้อทั้งหมด เนื่องจากข้อมูลไม่เพียงพอ นอกจากนี้ยังมีข้อจำกัดบางประการเกี่ยวกับค่าประมาณพารามิเตอร์ของข้อด้วย

ค่าประมาณความสามารถ ($\hat{\theta}_k$) ทั้งหมดที่ได้มีจุดตั้งต้น (origin) และหน่วยการวัด (unit of measurement) ไม่แน่นอน (arbitrary) (ผจงจิต อินทสุวรรณ, 2528, หน้า 50-54 ตอน 3.5) จึงมีการแปลงเชิงเส้นตรงเพื่อป้องกันเหตุการณ์ที่ค่า $\hat{\theta}$ สูงหรือต่ำเกินไปและค่าประมาณอำนาจจำแนก (a_1) และความยาก (b_1) ก็จะถูกปรับตามไปด้วย การแปลงเชิงเส้นตรงนี้ทำให้ $\hat{\theta}_i$'s มีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานอยู่ในช่วงที่กำหนด หรือ ± 3.0

ค่าอำนาจจำแนก (a_1) บางค่าอาจประมาณได้ไม่ดีและอาจมีค่าสูงเกินไป จึงมีการกำหนดขอบเขตของ a_1

การประมาณค่า c_1 กระทำได้ไม่ดีนัก จึงต้องมีข้อจำกัดมาก ข้อจำกัดนี้ขึ้นอยู่กับหลายองค์ประกอบคือ ความคลาดเคลื่อนมาตรฐานของ $\hat{c}_1$ ค่ามัชฌิมของ $\hat{c}_1$ ทั้งหมด สัดส่วนการตอบถูกที่รับสำหรับการเว้นไม่ตอบ ค่าของ $(b_1 - 2/a_1)$ ค่าสูงสุด (.5) และค่าต่ำสุด (0) ของ c_1

การใช้โปรแกรมคอมพิวเตอร์ LOGIST มีรายละเอียดที่ผู้ให้ต้องศึกษาอีกมาก โปรแกรมนี้เป็นที่รู้จักกันดี และใช้กันอย่างกว้างขวางมากที่สุดในกลุ่มนักวิจัยและผู้เชี่ยวชาญในการวัดทางการศึกษา และจิตวิทยา โปรแกรมชุดล่าสุด มีบริการอยู่

ที่ศูนย์คอมพิวเตอร์จุฬาลงกรณ์มหาวิทยาลัย (สุพัฒน์ สุกมลสันต์ 2530) งานวิจัยที่ทำในประเทศไทยที่เกี่ยวข้องกับการใช้LOGISTส่วนใหญ่เป็นงานวิจัยระดับวิทยานิพนธ์ ดังเช่น จิราพร ไกรสรศิริเวช (2529) ชรรค์ชัย คงเสน่ห์ (2530) เรวดี อินทสระ (2530) สุนันท์ ศลโกสม (2530) เสวี บรมชาลิตโรจน์ (2531) ชีรศักดิ์ อรุณานนท์ (2531) และ บุญมี พันธุ์ไทย (2531)

วิธีการประมาณค่าพารามิเตอร์สำหรับโมเดลของราช

สำหรับโมเดลโลจิสติกที่มีพารามิเตอร์ของข้อเพียงตัวเดียว (โมเดล 1PL) หรือโมเดลของราช (Rasch model) (Rasch, 1960) หรือโมเดลโลจิสติกอย่างง่าย (simple logistic model--SLM) (Andrich, 1988) มีข้อตกลงเบื้องต้นเพิ่มขึ้นจากโมเดลที่มีพารามิเตอร์ 3 ตัว นั่นคือ สมมติว่าในการตอบแต่ละข้อนั้น ผู้ตอบไม่มีการเดา ($c_i = 0$) และอำนาจจำแนกของแต่ละข้อ (a_i) มีค่าเท่ากันหมดทุกข้อ และเพื่อความสะดวกกำหนดให้เป็น 1 ดังนั้นจึงมีพารามิเตอร์อยู่เพียงตัวเดียวคือความยาก (b_i) ที่จะต้องประมาณค่า

ในโมเดล 1PL นี้ จำนวนคำตอบถูกของผู้สอบคนหนึ่งเป็นสถิติที่เพียงพอสำหรับการประมาณค่าความสามารถ (θ) ของเขา ข้อมูลอื่นใดนอกเหนือจากนี้ ย่อมไม่ได้ช่วยในการประมาณค่าความสามารถของเขาให้ดีขึ้นเลย จากการที่จำนวนคำตอบถูกเป็นสถิติที่เพียงพอ จึงสามารถพิจารณาผู้สอบทุกคนที่มีคะแนนเท่ากันว่าเขาเหล่านั้นเท่าเทียมกัน (equivalent) ในความสามารถ ดังนั้นจึงประมาณค่าความสามารถสำหรับแต่ละกลุ่มผู้สอบที่มีความสามารถเท่าเทียมกัน และค่าความยากของแต่ละข้อ ดูวิธีการประมาณอย่างง่ายในราช (Rasch, 1960) ค่าประมาณความสามารถมีลักษณะคงเส้นคงวา (consistent)

นอกจากนี้การประมาณค่าในโมเดลโลจิสติกอย่างง่ายนี้ ยังสามารถกระทำ
ได้โดยวิธีไล่คัลลัสสูงสุดแบบมีเงื่อนไขหรือแบบไม่มีเงื่อนไข ซึ่งให้ค่าประมาณที่ดี
(Wright and Panchapakesan, 1969; Andersen, 1970; Wright and Bell,
1977; and Wright and Douglas, 1977)

โปรแกรมคอมพิวเตอร์ชื่อ BICAL ได้รับการพัฒนาขึ้นเพื่อให้ประมาณ
ค่าพารามิเตอร์ของข้อ (ความยาก) และของผู้สอบ (ความสามารถ) (Wright
and Mead, 1978) วิธีการประมาณค่าพารามิเตอร์ที่ใช้ในโปรแกรมนี้นี้ 2 วิธีคือ
วิธีไล่คัลลัสสูงสุด แบบไม่มีเงื่อนไข (UCON) และวิธีคำนวณโดยประมาณ (PROX)
วิธีแรกมีกระบวนการในการประมาณค่าความยากของข้อ และความสามารถของผู้สอบ
ร่วมกัน จากผลการตอบของผู้สอบ ซึ่งอยู่ในรูปของเมทริกซ์ขนาด $n \times N$
(n คือจำนวนข้อ N คือจำนวนคน) ของคะแนน 0-1 เช่นเดียวกับที่ใช้ในโปรแกรม
LOGIST กระบวนการประมาณค่าเป็นไล่คัลลัสสูงสุดซึ่งใช้การประมาณค่าซ้ำ ๆ
เช่นกัน ส่วนวิธี PROX ตั้งข้อสมมติว่าความสามารถมีการแจกแจงเป็นแบบปกติ และ
ใช้เวลาในการคำนวณน้อยกว่า

ตัวอย่างงานวิจัยในประเทศไทย ที่ใช้โปรแกรม BICAL ในการประมาณ
ค่าพารามิเตอร์เช่นงานของ บัญชา แสงทวี (2530) อภรณ์ กาญจนกิจโสภณ (2531)
และสุรัตน์ ทวีบุญจันทร์ (2531)

เมื่อข้อมูลมีขนาดเล็ก วิธีการประมาณค่าพารามิเตอร์อาจกระทำได้โดยวิธี
คำนวณด้วยมือ (Wright, 1977) วิธีนี้สามารถใช้เครื่องคิดเลขช่วยคำนวณได้แต่จะใช้เวลา
นาน การใช้คอมพิวเตอร์ขนาดเล็กจะช่วยลดภาระในการคำนวณได้มาก

เพื่อประโยชน์แก่ครูในการวิเคราะห์ผลการสอบของนักเรียนของตน

มาสเตอร์ (Masters, 1984) ได้พัฒนาโปรแกรมชื่อ DICOT โดยเน้นที่การพิมพ์ผลการวิเคราะห์ให้อยู่ในรูปที่อ่านง่าย และมีผลต่าง ๆ ที่เป็นประโยชน์ต่อครูโดยตรง นอกจากนี้ยังแปลงค่าประมาณความสามารถและความยากจากมาตราโลจิต (logits) ให้เป็นวัต (WATS) เพื่อให้ค่าเฉลี่ยของความยากเป็น 50 และ ให้ความสามารถและความยากมีค่าจาก 0 ถึง 100

ในการศึกษาครั้งนี้ใช้โปรแกรม BICAL และกำหนดวิธี UCON ในการประมาณค่าพารามิเตอร์ โปรแกรม BICAL มีบริการที่ศูนย์คอมพิวเตอร์กระทรวงศึกษาธิการ

บทที่ 2

ระเบียบวิธี

แบบทดสอบ

แบบทดสอบที่นำมาวิเคราะห์ผลการตอบในครั้งนี้เป็นแบบทดสอบวัดผลสัมฤทธิ์ วิชาคณิตศาสตร์ระดับชั้นประถมศึกษาปีที่ 6 ในกรุงเทพมหานคร ผู้สร้างแบบทดสอบเป็นครูคณิตศาสตร์ ในโรงเรียนจำนวน 10 โรงเรียน ในเขตบางกะปิ และให้ทดสอบปลายปี วัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ของนักเรียนชายหญิงในกลุ่มโรงเรียนดังกล่าว ปีการศึกษา 2529 ลักษณะแบบทดสอบเป็นแบบเลือกตอบมีตัวเลือก 4 ตัวเลือก ฉบับที่ใช้เป็นฉบับส่งกับทาง คณิตศาสตร์มีจำนวน 30 ข้อ และฉบับโจทย์ปัญหาจำนวน 30 ข้อ รวมทั้งหมดจำนวน 60 ข้อ ในการวิเคราะห์ครั้งนี้ถือเป็นแบบทดสอบฉบับเดียวซึ่งมี 60 ข้อ นอกจากนี้การเฉลยคำตอบ ข้อ 14 และ 60 กำหนดให้คลาดเคลื่อนเพื่อการตรวจสอบผลการวิเคราะห์

กลุ่มตัวอย่าง

ผู้ตอบแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์เป็นนักเรียนชั้น ประถมปีที่ 6 ในเขตบางกะปิ ของกรุงเทพมหานคร จำนวนทั้งหมด 1,502 คน ในโรงเรียน 10 โรงเรียน ดังตารางแสดงจำนวนนักเรียนในแต่ละโรงเรียนตาราง 1

วิธีการตรวจสอบความเป็นมิติเดียว

ข้อมูลผลการตอบข้อคำถามที่จะนำมาประมาณค่าพารามิเตอร์ตามโมเดล โลกจริงตึก ต้องแสดงให้เห็นว่ามีมิติเดียว ในการศึกษาครั้งนี้ได้ดำเนินการตรวจสอบความเป็นมิติเดียวโดยมีขั้นตอนดังนี้

1. ตรวจให้คะแนนข้อคำถามแต่ละข้อ โดยให้เกณฑ์ตอบถูกให้ 1 คะแนน ตอบผิดให้ 0 คะแนน ถ้าไม่ตอบถือเป็นไม่มีข้อมูล (missing)

ตาราง 1 แสดงจำนวนนักเรียนในกลุ่มตัวอย่างจำแนกตามโรงเรียน

โรงเรียน	จำนวนนักเรียน	โรงเรียน	จำนวนนักเรียน
วัดบางเตย	163	พิชัยพัฒนา	66
วัดลาดปลาเค้า	218	คันทนาสาว	75
วัดลาดพร้าว	89	วัดบึงทองหลาง	274
วัดศรีบุญเรือง	74	บ้านบางกะปิ	409
วัดสุเหร่าชีรอ	58	ลำสาลี	76
		รวม	1,502

2. จากคะแนนราชทัณฑ์สำหรับผู้สอบแต่ละคนตามที่ตรวจได้ในข้อ 1 ซึ่งเป็นเมทริกซ์ขนาด 1502×60 นำมาใช้ในการคำนวณสัมประสิทธิ์สหสัมพันธ์ระหว่างข้อสหสัมพันธ์ที่ใช้คือเทตระคลริค (tetrachoric correlation) การคำนวณนี้ใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป BMDP4F (Dixon & Brown, 1979) ผลที่ได้ในขั้นนี้เป็นเมทริกซ์สหสัมพันธ์ขนาด 60×60

3. คำนวณค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณกำลังสอง (square multiple correlation--SMC) แล้วใช้ SMC แทนค่า 1 ที่เส้นทแยงมุมของเมทริกซ์สหสัมพันธ์ที่ได้ในขั้นที่ 2 นำเมทริกซ์สหสัมพันธ์ที่มี SMC ที่เส้นทแยงมุมนี้มาคำนวณค่าไอเกน (eigen values) และค่าร้อยละของความแปรปรวนที่อธิบายได้ (percent variance

explained) โดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป SAS (SAS, 1985)

หมายเหตุ ผู้ใช้สามารถระบุให้ SAS ให้แมทริกซ์สหสัมพันธ์ที่มี SMC ที่
เส้นทะแยงมุมในการคำนวณค่าไอเกนได้ โดยโปรแกรมนี้จะคำนวณค่า SMC และค่าไอเกน
ได้ในการใช้คอมพิวเตอร์ทำงานครั้งเดียว

การประมาณค่าพารามิเตอร์

ตรวจให้คะแนนการตอบข้อคำถามของนักเรียนแต่ละคนเป็น 4 แบบคือ
ตอบถูกให้คะแนน 1 ตอบผิดให้ 0 เว้นไม่ตอบให้ 2 และทำไม่ถึงให้ 3 (Wood, et
al., 1976; and Wingersky, et al., 1982) แล้วใช้แมทริกซ์ของคะแนน
ราชทัณฑ์สำหรับผู้ตอบแต่ละคนที่ตรวจแล้ว ซึ่งเป็นแมทริกซ์ขนาด 1502 x 60 เป็นข้อมูล
นำเข้าสำหรับโปรแกรม LOGIST เพื่อประมาณค่าพารามิเตอร์ของข้อภาษาได้โมเดล 3PL
พารามิเตอร์ทั้งสามตัวคือ อำนาจจำแนก (a) ความยาก (b) และ แอซิมโทคล้าง
หรือการเดา (c)

สำหรับการวิเคราะห์ราชทัณฑ์โดยใช้โปรแกรม BICAL นั้นข้อมูลนำเข้ามา
จากการตรวจให้คะแนนการตอบข้อคำถามเป็น 2 แบบคือ ตอบถูกให้คะแนน 1 นอกนั้น
ให้ 0 (Wright & Mead, 1978; and Wright, et al., 1980) พารามิเตอร์
ที่ประมาณมีเพียงตัวเดียวคือความยาก (b) ซึ่งเป็นไปตามโมเดลของราช และกำหนด
วิธีการประมาณเป็นไลค์ลีสต์สูงสุดแบบไม่มีเงื่อนไข (UCON)

ส่วนการประมาณค่าพารามิเตอร์ความง่าย (p) และอำนาจจำแนก
(r_{11}) ตามโมเดลที่ใช้ยกตั้งเดิมนั้นใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป SPSS*
(Nie, 1983) คำสั่ง FREQUENCIES ให้ค่าความง่ายของข้อซึ่งคำนวณมาจากสัดส่วน
คนตอบถูกจากคนที่ทำข้อนั้น และคำสั่ง RELIABILITY ให้ค่าสหสัมพันธ์พอยท์ไบชีเรียล

ระหว่างข้อหนึ่ง ๆ กับคะแนนรวมของทุกข้อในฉบับนั้นยกเว้นข้อที่นำมาหาความสัมพันธ์

นอกจากนั้นยังแสดงข้อมูลละเอียดตามแนวทางการวิเคราะห์รายข้อของ

หน่วยบริการทดสอบทางการศึกษา (Educational Testing Service--ETS)

ประเทศสหรัฐอเมริกา (Wainer, 1989) ลำดับขั้นในการวิเคราะห์รายข้อมีดังนี้

1. แบ่งนักเรียนทั้งหมดออกเป็น 5 กลุ่มเท่า ๆ กันโดยประมาณตามคะแนนรวม กลุ่มแรกเป็นกลุ่มที่ได้คะแนนช่วงต่ำสุด และกลุ่มที่ 2 ถึงกลุ่มที่ 4 มีคะแนนสูงขึ้นมาตามลำดับจนถึงกลุ่มที่ 5 เป็นกลุ่มที่มีคะแนนสูงสุด (ขั้นนี้ดัดแปลงจากวิธีการของ ETS เนื่องจากวิธีการของ ETS คิดเฉพาะนักเรียนที่ทำถึงข้อนี้)

2. สำหรับแต่ละกลุ่มนักเรียน นับจำนวนความถี่ของนักเรียนที่ตอบตัวเลือกต่าง ๆ

4 ตัวเลือก

ผลจากข้อ 1 และข้อ 2 จึงได้เป็นตารางคอนติเนนต์ที่มี 2 มิติ (ใช้ CROSSTABS ใน SPSS^x (Nie, 1983) ตัวเลขภายในตารางเป็นความถี่หรือจำนวนนักเรียนที่ได้คะแนนอยู่ในกลุ่มใดกลุ่มหนึ่งของช่วงคะแนน และตอบตัวเลือกตัวใดตัวหนึ่งใน 4 ตัวเลือก

3. ใช้คำสั่ง BREAKDOWN ใน SPSS^x (Nie, 1983) คำนวณค่าต่อไปนี้

3.1 ค่าคะแนนเฉลี่ยทั้งฉบับของนักเรียนที่เลือกตอบแต่ละตัวเลือก (แทนด้วย M_A , M_B , M_C , และ M_D ซึ่งหมายถึงค่าเฉลี่ยของตัวเลือก 1, 2, 3, และ 4 ตามลำดับ)

3.2 ค่าคะแนนเฉลี่ยทั้งฉบับของนักเรียนที่ทำข้อนี้ (แทนด้วย M_{total})

3.3 สัดส่วนของกลุ่มตัวอย่างที่ทำถึงข้อนี้ (แทนด้วย P_{total})

3.4 สัดส่วนของผู้สอบในกลุ่มตัวอย่างที่เลือกตัวเลือกที่ถูกจากคนที่ทำ

ถึงข้อนี้ (แทนด้วย P_i) นี้เป็นดัชนีวัดความง่ายของข้อ

ผลการตอบข้อหนึ่ง ๆ ที่เกินไปตามโครงสร้างที่กลมรับได้เป็นดังนี้

1. เมื่อคะแนนรวมสูงขึ้น จำนวน (หรือสัดส่วน) คนที่ตอบข้อเลือกที่ถูกควรสูงขึ้นด้วย
2. เมื่อคะแนนรวมสูงขึ้น จำนวน (หรือสัดส่วน) คนที่เลือกข้อเลือกที่เป็นตัวลวง
ทั้งหลายควรลดลง

การหาความสัมพันธ์ระหว่างค่าประมาณความยากจากโมเดลทั้งสาม

ค่าประมาณความยากภายใต้โมเดล 3PL(b) โมเดล 1PL(b) และโมเดล
ดั้งเดิม(p) นำมาจัดลำดับโดยเรียงค่า b จากต่ำไปสูง ส่วนค่า p เรียงจากสูงไปต่ำ
แล้วคำนวณความสัมพันธ์เชิงอันดับที่ละคู่ โดยใช้สูตร สเปียร์แมน ไร (Spearman Rho)
(Hinkle, et al., 1979, p.101-104)

บทที่ 3

ผลการวิเคราะห์ข้อมูล

การรายงานผลการวิเคราะห์ข้อมูล ในส่วนแรกเป็นผลการวิเคราะห์เนื้อหา และพฤติกรรมที่แบบทดสอบฉบับนี้วัด ส่วนที่สองเป็นผลการตรวจสอบความเป็นมิติเดียวของคะแนนจากแบบทดสอบ ส่วนที่สามเป็นการรายงานข้อมูลพื้นฐานของคะแนนรวมจากแบบทดสอบ ส่วนที่สี่เป็นการรายงานผลการวิเคราะห์รายข้อซึ่งมีลำดับดังนี้

(1) ผลการประมาณค่าพารามิเตอร์ภายใต้โมเดล 3PL จากโปรแกรม LOGIST (2) ผลการประมาณค่าพารามิเตอร์ภายใต้โมเดล 1PL หรือโมเดลของราทจากโปรแกรม BICAL และ (3) ผลการประมาณค่าพารามิเตอร์ภายใต้โมเดลที่ใช้กันอยู่ดั้งเดิม โดยอาศัยโปรแกรมสำเร็จรูป SPSS และ ส่วนสุดท้ายเป็นการคำนวณความสัมพันธ์เชิงอันดับระหว่างความยากของข้อที่ประมาณภายใต้โมเดลโลจิสติกและโมเดลดั้งเดิม

ผลการวิเคราะห์เนื้อหาและพฤติกรรมที่แบบทดสอบฉบับนี้วัด

เนื้อหาวิชาคณิตศาสตร์สำหรับชั้นประถมศึกษาปีที่ 6 ตามที่กำหนดไว้ในหนังสือเรียนคณิตศาสตร์ (สสวท, 2532) แบ่งออกเป็น 19 บท ส่วนพฤติกรรมที่หาค่าถามแต่ละข้อในแบบทดสอบฉบับนี้วัดค่ามาตรฐานจากกฎการจำแนกของบลูม (Bloom, et al., 1971) ซึ่งแบ่งการวัดผลสัมฤทธิ์ทางการเรียนด้านความคิด (cognitive domain) ไว้ 6 พฤติกรรม คือ ความรู้-ความจำ ความเข้าใจ การนำไปใช้ การวิเคราะห์ การสังเคราะห์ และการประเมินค่า ผลการวิเคราะห์เนื้อหาและพฤติกรรมปรากฏในตาราง 2 และภาคผนวก ก

ผลการตรวจสอบความเป็นมิติเดียวของคะแนนจากแบบทดสอบ

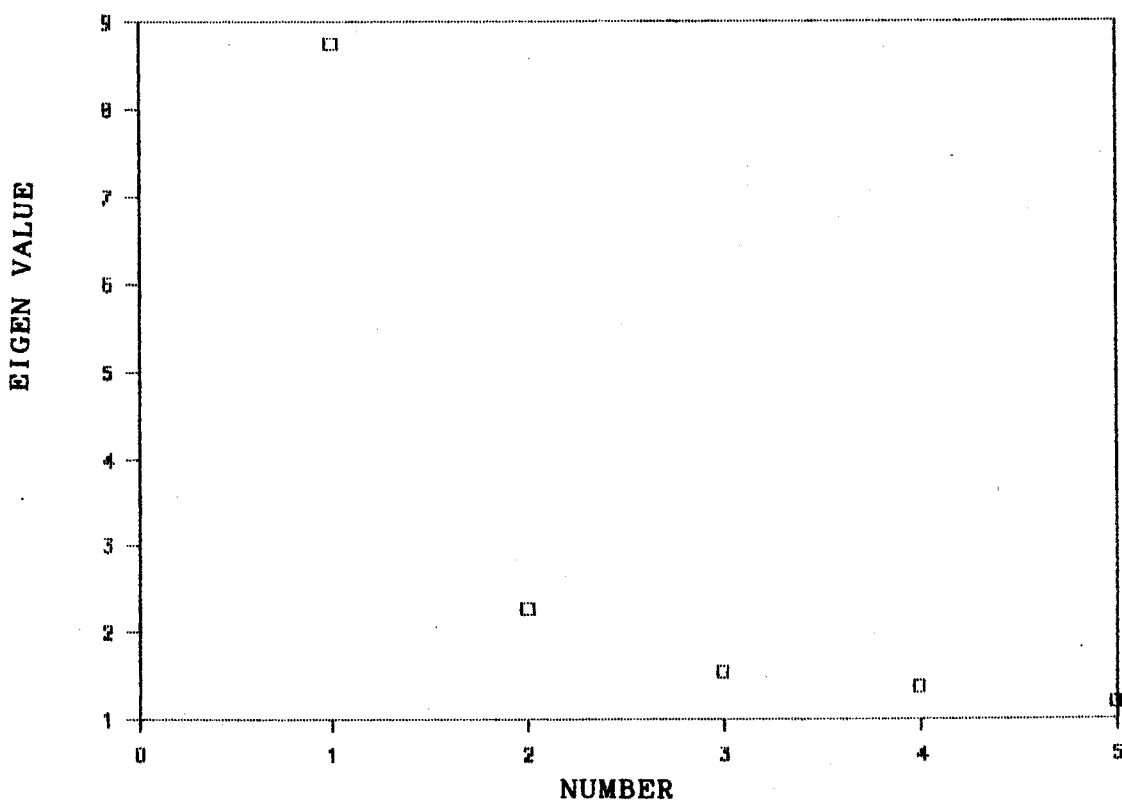
ผลการคำนวณค่าไคเอนจากเมทริกซ์สหสัมพันธ์เทตระคอร์ริเกอระหว่างข้อที่มี SMC ที่เส้นทะแยงมุม ปรากฏว่าค่าไคเอนค่าแรกคือ 8.75 ซึ่งอธิบายความแปรปรวน

ตาราง 2 แสดงจำนวนหัวข้อมาตรการสวัสดิภาพสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์

ทั้งประเภทที่ 6 จำนวนตามเนื้อหาและพฤติกรรมต่าง ๆ ที่วัด

เนื้อหา	พฤติกรรม				รวม
	ความจำ	ความเข้าใจ	การนำไปใช้	การวิเคราะห์	
1. จำนวนและตัวเลข	5	-	-	-	5
2. สมการ	1	3	3	-	7
3. ตัวประกอบของจำนวนนับ	2	-	-	-	2
4. ความเท่ากันทุกประการ	-	-	-	1	1
5. รูปสมมาตร	-	-	-	2	2
6. มุมและส่วนของเส้นตรง	2	-	-	-	2
7. เศษส่วน	-	-	3	-	3
8. ทศนิยม	-	-	1	-	1
9. ทศนิยม	-	-	1	-	1
10. เส้นขนาน	-	-	2	-	2
11. รูปสามเหลี่ยม	-	-	-	-	-
12. บทประยุกต์	-	-	7	-	7
13. ทิศและแผนผัง	3	5	1	-	9
14. รูปสี่เหลี่ยม	-	-	1	-	1
15. ความยาวรอบรูปและพื้นที่					
ของรูปเรขาคณิต	1	1	1	-	6
16. ความยาวรอบรูปและพื้นที่					
ของรูปวงกลม	-	-	2	-	2
17. รูปทรงและปริมาตร	-	-	2	2	4
18. คู่ขนาน	-	-	1	-	1
19. แผนภูมิและกราฟ	-	3	1	-	4
รวม	14	12	29	5	60

ของคะแนนรวมได้ร้อยละ 38.48 และค่าไอเกนค่าต่อ ๆ ไปมีค่าต่ำมากเมื่อเทียบกับค่าแรก ฉะนั้นแฟคเตอร์ที่ 1 นี้ จึงอธิบายความแปรปรวนส่วนใหญ่ของคะแนนจากแบบทดสอบ จึงกล่าวได้ว่าคะแนนผลการตอบข้อคำถามในแบบทดสอบฉบับนี้มีมิติเดียว ยอมรับได้ว่าสามารถใช้วิเคราะห์คุณภาพของข้อคำถามได้ (Reckase.1979) ภาพ 2 แสดงขนาดในเชิงเปรียบเทียบของค่าไอเกน 5 ค่าแรก พร้อมทั้งค่าร้อยละของความแปรปรวนของคะแนนรวมที่สามารถอธิบายได้ด้วยแฟคเตอร์แต่ละแฟคเตอร์



ค่าไอเกน	8.75	2.27	1.54	1.37	1.19
เปอร์เซ็นต์ความ					
แปรปรวน ที่อธิบายได้	38.48	9.98	6.78	6.04	5.24

ภาพ 2 แสดงค่าไอเกนห้าค่าแรกและค่าร้อยละของความแปรปรวน
ที่แต่ละแฟคเตอร์อธิบายคะแนนรวมได้

ข้อมูลพื้นฐานของคะแนนรวมจากแบบทดสอบ

คะแนนรวม (60 ข้อ) ผลการสอบวิชาคณิตศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 1502 คน คำนวณค่าสัมประสิทธิ์แอลฟา (alpha coefficient) ได้

.8215 ลักษณะการกระจายของกลุ่มผู้สอบมีดังนี้ คะแนนต่ำสุด 11 คะแนนสูงสุด 57 (จากคะแนนเต็ม 60) ค่าเฉลี่ยอยู่ที่ 32.472 โดยมีส่วนเบี่ยงเบนมาตรฐาน 8.063 ความเบ้ .176 (ค่าใกล้ 0 แสดงว่าการกระจายของคะแนนใกล้เคียงลักษณะสมมาตร)

การแจกแจงความถี่ของคะแนนแสดงไว้ในตาราง 3 ส่วนภาพ 3 แสดงรูปหลายเหลี่ยมแห่งความถี่ของคะแนนดังนี้

ผลการประมาณค่าพารามิเตอร์ภายใต้โมเดลโลจิสติกที่มีพารามิเตอร์ 3 ตัว (โมเดล 3PL)

ผลการตรวจสอบคะแนนรวมแสดงว่าไม่มีผู้สอบคนใดได้คะแนน 0 หรือ คะแนนเต็ม ดังนั้นในการประมาณค่าพารามิเตอร์ของข้อโดยใช้โปรแกรม LOGIST ครั้งนี้จึงใช้ผลการตอบของผู้สอบทุกคน

การประมาณค่าพารามิเตอร์ของข้อคำถามทั้ง 60 ข้อโดยใช้ LOGIST ให้ผลไม่สมบูรณ์ เนื่องจากข้อที่ 14 มีค่าดีเทอร์มิแนนต์ (determinant) ของแมทริกซ์ของอนุพันธ์ที่สอง (second derivatives) เป็น 0 ดังนั้นจึงประมาณค่าพารามิเตอร์โดยใช้ข้อมูลจาก 59 ข้อ โดยตัดข้อที่ 14 ออก ค่าประมาณพารามิเตอร์ของข้อคือ อำนาจจำแนก (a) ความยาก (b) และแนวโน้มถ่วงหรือการเดา (c) ของข้อคำถาม 59 ข้อ ปรากฏในตาราง 4

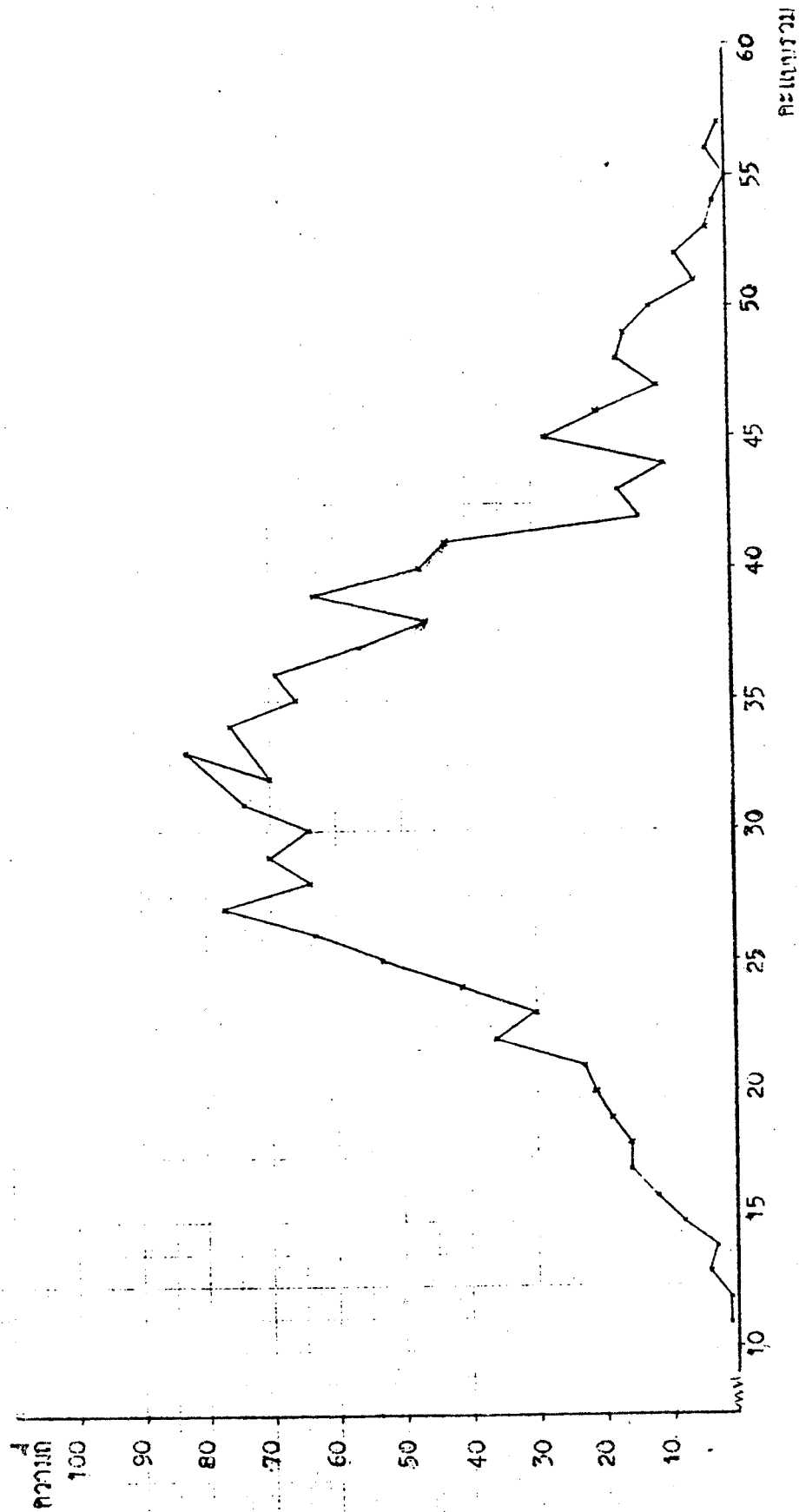
ตาราง 3 แสดงการแจกแจงความถี่ของคะแนนรวมจากแบบทดสอบวัดผลสัมฤทธิ์

วิชาคณิตศาสตร์ที่มี 60 ข้อ ของนักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 1502 คน

คะแนน	ความถี่	ร้อยละ	ร้อยละสะสม
11	1	.1	.1
12	1	.1	.1
13	4	.3	.4
14	3	.2	.6
15	8	.5	1.1
16	12	.8	1.9
17	16	1.1	3.0
18	16	1.1	4.1
19	19	1.3	5.3
20	21	1.4	6.7
21	23	1.5	8.3
22	36	2.4	10.7
23	30	2.0	12.6
24	41	2.7	15.4
25	53	3.5	18.9
26	63	4.2	23.1
27	77	5.1	28.2
28	64	4.3	32.5
29	70	4.7	37.2
30	64	4.3	41.4
31	74	4.9	46.3
32	70	4.7	51.0
33	83	5.5	56.5
34	76	5.1	61.6

ตาราง 3 (ต่อ)

คะแนน	ความถี่	ร้อยละ	ร้อยละสะสม
35	66	4.4	66.0
36	69	4.6	70.6
37	56	3.7	74.3
38	46	3.1	77.4
39	63	4.2	81.6
40	47	3.1	84.7
41	33	2.2	86.9
42	24	1.6	88.5
43	27	1.8	90.3
44	20	1.3	91.6
45	28	1.9	93.5
46	20	1.3	94.8
47	11	.7	95.5
48	17	1.1	96.6
49	16	1.1	97.7
50	12	.8	98.5
51	5	.3	98.9
52	8	.5	99.4
53	3	.2	99.6
54	2	.1	99.7
55	0	0	99.7
56	3	.2	99.9
57	1	.1	100.0
1502		100.0	



ภาพ 3 แสดงการเปลี่ยนแปลงความเข้มของคลื่นโพลาไรซ์

ตาราง 4 แสดงค่าประมาณพารามิเตอร์ ถ่านาจจำแนก (a) ความยาก (b) และ
การเดา (c) ของข้อคำถาม 59 ที่อยู่ในแบบทดสอบวัดผลสัมฤทธิ์วิชา
คณิตศาสตร์จากผลการตอบของนักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 1502 คน

ลำดับที่	ข้อที่	a	b	c
1	1	0.87721	-1.99850	*0.16458
2	2	0.85302	-0.22832	0.12765
3	3	0.52630	-2.91642	*0.16458
4	4	0.78705	1.12825	0.24284
5	5	0.90811	0.91818	0.22854
6	6	0.57567	2.18404	0.09891
7	7	0.50237	-1.90385	*0.16458
8	8	0.46081	-1.62584	*0.16458
9	9	0.36461	-2.57058	*0.16458
10	10	0.70390	0.65849	0.25327
11	11	0.64031	-0.64829	*0.16458
12	12	0.40680	-2.00256	*0.16458
13	13	0.74110	2.08316	0.17968
14	15	0.22787	-0.15092	*0.16458

ตาราง 4 (ต่อ)

ลำดับที่	ข้อที่	a	b	c
15	16	0.30412	-0.01947	*0.16458
16	17	0.44398	-3.74895	*0.16458
17	18	0.42249	-1.98946	*0.16458
18	19	0.50864	-0.97543	*0.16458
19	20	0.93031	0.84568	0.33198
20	21	0.90832	0.82439	0.39662
21	22	1.08467	2.18452	0.20653
22	23	0.92650	0.16590	0.26362
23	24	0.47568	-0.50332	*0.16458
24	25	0.55398	2.86998	0.20316
25	26	0.72555	0.22893	*0.16458
26	27	0.61818	0.96019	0.17152
27	28	0.76626	1.57230	0.17313
28	29	0.73375	-0.76167	*0.16458
29	30	0.14713	-0.03812	*0.16458
30	31	0.97039	-0.02710	0.27544
31	32	0.35199	-0.37371	*0.16458

ตาราง 4 (ต่อ)

ลำดับที่	ปีที่	a	b	c
32	33	1.29021	1.56949	0.33350
33	34	2.00000	2.62323	0.30625
34	35	0.38967	2.83308	0.30078
35	36	0.78422	1.52564	0.23601
36	37	0.89450	1.07376	0.15182
37	38	1.04018	1.59799	0.18897
38	39	0.84292	1.79656	0.09357
39	40	0.61092	-0.01400	*0.16458
40	41	0.26629	3.91101	0.04184
41	42	0.92747	1.68130	0.18828
42	43	0.90395	0.78156	0.23537
43	44	0.97796	0.92835	0.30208
44	45	1.09545	-0.27404	0.13713
45	46	0.46555	1.07285	*0.16458
46	47	1.06232	-1.21385	*0.16458
47	48	1.17140	-0.22199	0.28751
48	49	0.79152	0.16032	0.07632

ตาราง 4 (ต่อ)

ลำดับที่	ข้อที่	a	b	c
49	50	0.53881	-0.12606	*0.16458
50	51	0.99844	1.31108	0.27857
51	52	0.92540	-0.59516	*0.16458
52	53	0.80574	0.23955	0.08902
53	54	0.48850	-0.77355	*0.16458
54	55	0.60167	0.08804	*0.16458
55	56	0.89364	0.30571	0.24614
56	57	0.45874	-0.28361	*0.16458
57	58	0.90052	2.90887	0.15756
58	59	0.57651	-1.20593	*0.16458
59	60	0.17156	1.33699	*0.16458

* หมายถึง โปรแกรมกำหนดค่า c คงที่เนื่องจากค่าที่ได้ไม่ดี

ได้แสดงลักษณะของหาค่าถามแต่ละข้อ ซึ่งมีค่าประมาณพารามิเตอร์ตาม

ในตาราง 4 ปรากฏในภาคผนวก ก

ผลในตาราง 4 แสดงว่าค่าประมาณพารามิเตอร์ อำนาจจำแนก (a) มีช่วงจาก .1471 ถึง 2.0000 ข้อคำถามที่มีอำนาจจำแนกต่ำกว่า .2000 ($a < .2000$) มี 2 ข้อคือ ข้อ 30 และ 60 ส่วนข้อที่มีค่าอำนาจจำแนกต่ำกว่า .3000 ($a < .3000$) มี 4 ข้อคือ ข้อ 15, 30, 41, และ 60 ข้อคำถามเหล่านี้มีโด่ง (icc) ค่อนข้าง แบนราบ (ดูภาคผนวก ข)

ค่าประมาณความยาก (b) มีช่วงจาก -3.7490 ถึง 3.9110

ค่าประมาณแอซิมโทติกล่างหรือการเดา (c) ต่ำสุดเท่ากับ .0418 สูงสุดเท่ากับ .3966 และมีข้อคำถามจำนวน 25 ข้อ ที่โปรแกรมต้องกำหนดค่าคงที่เนื่องจาก ค่าที่ได้ไม่ดี หรือ (b-2/a) มีค่าน้อยกว่า -2.5 (Wood, et al., 1976, p.17-18)

ผลการประมาณค่าพารามิเตอร์ในโมเดลของราช

คะแนนที่นำมาวิเคราะห์โดยใช้โปรแกรม BICAL ในครั้งนี้ เป็นผลการสอบ ของนักเรียนจำนวน 1,485 คน นั่นคือมีการคัดนักเรียนออก 17 คน เนื่องจากนักเรียน เหล่านี้มีคะแนนรวมเท่ากับหรือน้อยกว่าระดับการเดา (คะแนน 15 หรือต่ำกว่า ดูตาราง 2 ค่าประมาณพารามิเตอร์ความยากของข้อ (b) และค่าสถิติที (t-test) สำหรับทดสอบ ความพอเหมาะของข้อคำถามทั้ง 60 ข้อ ปรากฏในตาราง 5

จากการพิจารณาค่าสถิติที (t) สำหรับทดสอบความพอเหมาะ ปรากฏว่า ค่าสถิติดังกล่าวสำหรับข้อ 14 สูงมาก ดังนั้นจึงวิเคราะห์ข้อมูลโดยใช้โปรแกรม BICAL อีกครั้งหนึ่งโดยตัดข้อ 14 ออกจากการวิเคราะห์ ในครั้งนี้จึงเป็นการวิเคราะห์ 59 ข้อ ค่าประมาณครั้งหลังแสดงในตาราง 6 ในตารางนี้ค่าประมาณความยากมีช่วงจาก -2.68 ถึง 1.90 และมีค่าเฉลี่ยเป็น 0 ส่วนเบี่ยงเบนมาตรฐานเป็น 1 ตามที่กำหนดไว้ในโปรแกรม

ตาราง 5 แสดงค่าประมาณพารามิเตอร์ความยาก (b) ความคลาดเคลื่อนมาตรฐาน

(Std error) และการทดสอบที่ของความพอเหมาะ (Fit t-test) ของข้อ
คำถามจำนวน 60 ข้อในแบบทดสอบวัดผลสัมฤทธิ์วิชาคณิตศาสตร์ ของนักเรียนชั้น
ประถมปีที่ 6 จำนวน 1,485 คน

ลำดับที่	ข้อที่	b	Std error	Fit t-test
1	1	-2.31	.10	-1.25
2	2	- .28	.06	-2.29
3	3	-2.40	.10	- .61
4	4	.57	.06	.00
5	5	.51	.06	- .68
6	6	1.59	.06	.09
7	7	-1.51	.07	- .60
8	8	-1.22	.07	- .52
9	9	-1.57	.07	- .09
10	10	.17	.05	- .06
11	11	- .68	.06	-1.06
12	12	-1.36	.07	- .09
13	13	1.32	.06	.50
14	14	1.48	.06	4.63
15	15	- .17	.06	1.75

ตาราง 5 (ต่อ)

ลำดับที่	ข้อที่	b	std error	Fit t-test
16	16	- .14	.06	1.39
17	17	-2.70	.11	- .46
18	18	-1.39	.07	- .30
19	19	- .83	.06	- .40
20	20	.16	.05	- .22
21	21	-.07	.06	.29
22	22	1.41	.06	1.00
23	23	- .20	.06	-1.51
24	24	- .47	.06	- .24
25	25	1.31	.06	1.71
26	26	.09	.05	-1.17
27	27	.59	.06	- .57
28	28	1.09	.06	- .08
29	29	- .85	.06	-1.42
30	30	- .14	.06	2.73
31	31	- .41	.06	-1.95
32	32	- .32	.06	.74

ตาราง 5 (ต่อ)

ลำดับที่	ข้อที่	b	std error	Fit t-test
33	33	.65	.06	.80
34	34	1.09	.06	2.70
35	35	.66	.06	2.36
36	36	.83	.06	.26
37	37	.91	.06	-1.47
38	38	1.20	.06	- .19
39	39	1.70	.07	- .89
40	40	-.13	.06	-1.07
41	41	1.83	.07	1.05
42	42	1.20	.06	.12
43	43	.40	.05	-1.29
44	44	.33	.05	- .37
45	45	- .38	.06	-3.07
46	46	.53	.06	.52
47	47	-1.56	.07	-2.21
48	48	- .65	.06	-2.17
49	49	.23	.05	-2.26

ตาราง 5 (ต่อ)

ลำดับที่	ข้อที่	b	std error	Fit t-test
50	50	- .22	.06	- .52
51	51	.65	.06	.23
52	52	- .76	.06	-2.46
53	53	.29	.05	-2.50
54	54	- .66	.06	- .49
55	55	- .06	.06	- .79
56	56	- .03	.05	-1.34
57	57	- .31	.06	- .18
58	58	1.86	.07	1.44
59	59	-1.09	.06	- .78
60	60	.22	.05	2.87

BICAL ค่าอำนาจจำแนกที่ใส่ไว้ในตาราง 6 มีประโยชน์ในการอธิบายเมื่อสถิติที่แสดงว่า
ข้อคำถามไม่เหมาะสมกับโมเดล แต่ไม่ได้เป็นพารามิเตอร์สำหรับบรรยายลักษณะของโด่ง
(icc) ในโมเดล

ตาราง 6 แสดงค่าประมาณพารามิเตอร์ความยาก (b) ความคลาดเคลื่อนมาตรฐาน

(Std error) อำนาจจำแนก (Disc index) และการทดสอบทีของความ

พลเหมาะสม (Fit t-test) ของข้อคำถามจำนวน 59 ข้อ ในแบบทดสอบวัด

ผลสัมฤทธิ์วิชาคณิตศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 1,485 คน

ลำดับที่	ข้อที่	b	Std error	Disc index	Fit t-test
1	1	-2.29	.10	1.59	-1.24
2	2	- .25	.06	1.48	-2.25
3	3	-2.38	.10	1.13	- .62
4	4	.60	.06	.97	.13
5	5	.54	.06	1.10	- .61
6	6	1.62	.07	.89	.26
7	7	-1.49	.07	1.13	- .59
8	8	-1.20	.07	1.10	- .48
9	9	-1.55	.07	.95	- .09
10	10	.19	.05	1.01	.00
11	11	- .66	.06	1.24	-1.03
12	12	-1.34	.07	.92	- .07
13	13	1.35	.06	.76	.63
14	15	- .15	.06	.57	1.86

ตาราง 6 (ต่อ)

ลำดับที่	ข้อที่	b	std error	Disc index	Fit t-test
15	16	- .11	.06	.64	1.51
16	17	-2.68	.11	.99	- .48
17	18	-1.37	.07	1.04	- .25
18	19	- .81	.06	1.01	- .34
19	20	.18	.05	1.03	- .18
20	21	- .05	.06	.92	.36
21	22	1.44	.06	.61	1.15
22	23	- .18	.06	1.33	-1.47
23	24	- .45	.06	1.00	- .16
24	25	1.35	.06	.53	1.85
25	26	.11	.05	1.21	-1.13
26	27	.61	.06	1.07	- .45
27	28	1.11	.06	.93	.06
28	29	- .82	.06	1.33	-1.43
29	30	- .11	.06	.33	2.90
30	31	- .38	.06	1.46	-1.90
31	32	- .30	.06	.77	.84

ตาราง 6 (ต่อ)

ลำดับที่	ข้อที่	b	std error	Disc index	Fit t-test
32	33	.68	.06	.74	.96
33	34	1.12	.06	.33	2.91
34	35	.69	.06	.45	2.51
35	36	.85	.06	.88	.38
36	37	.94	.06	1.32	-1.39
37	38	1.23	.06	.96	- .06
38	39	1.73	.07	1.12	- .74
39	40	- .10	.06	1.19	- .98
40	41	1.86	.07	.67	1.24
41	42	1.23	.06	.89	.24
42	43	.43	.05	1.26	-1.23
43	44	.36	.05	1.04	- .23
44	45	- .36	.06	1.71	-3.05
45	46	.56	.06	.81	.68
46	47	-1.54	.07	1.70	-2.20
47	48	- .63	.06	1.55	-2.13
48	49	.26	.05	1.50	-2.21

ตาราง 6 (ต่อ)

ลำดับที่	ข้อที่	b	std error	Disc index	Fit t-test
49	50	- .20	.06	1.10	- .45
50	51	.68	.06	.91	.34
51	52	- .74	.06	1.57	-2.47
52	53	.31	.05	1.48	-2.42
53	54	- .63	.06	1.08	- .47
54	55	- .04	.06	1.19	- .76
55	56	- .01	.06	1.31	-1.30
56	57	- .28	.06	1.00	- .10
57	58	1.90	.07	.37	1.58
58	59	-1.07	.06	1.15	- .77
59	60	.25	.05	.34	3.03

ได้แสดงลักษณะของข้อคำถามตามค่าประมาณพารามิเตอร์ความยากที่ปรากฏในตาราง 6 แสดงในภาคผนวก ค ในภาพที่แสดงได้ดังตามโมเดลของราชนันท์ การลงจุดตามค่าของสัดส่วนการตอบถูก (p) ไว้ด้วย แต่ละข้อมีจุดดังกล่าว 6 จุด ประกอบกันเป็นแนวของได้แสดงลักษณะของข้อคำถามที่ไม่ได้อาศัยโมเดลโลจิสติก ค่าของสัดส่วนการตอบถูก (p) ทั้ง 6 ค่าในแต่ละข้อนั้นคำนวณมาจากการแบ่งระดับ

ความสามารถ (หรือคะแนนดิบ) ออกเป็น 6 ช่วง แต่ละช่วงของความสามารถจะ
คำนวณค่า p ได้ 1 ค่า ค่า p ทั้ง 6 ค่า สำหรับหาค่าถามแต่ละข้อแสดงไว้ใน
ภาคผนวก ง

จากตาราง 6 พิจารณาคัดเลือกหาค่าถามโดยตรวจสอบสถิติที่สำหรับความ
เหมาะสมพบว่าข้อที่มีค่าที่อยู่นอกช่วง ± 2.576 [$\frac{z}{2} t_{\alpha} = 2.576$] คือข้อ 30, 34,
45 และ 60 หาค่าถามทั้ง 4 ข้อนี้ควรตัดออก เมื่อพิจารณาโค้งที่พล็อตจากสัดส่วน
การตอบถูกความคู่กับค่าสถิติที่พบว่า ข้อ 30, 34, และ 60 มีโค้งแบนราบ ส่วนข้อ 45
มีโค้งชันมาก แสดงให้เห็นว่าโมเดลไม่สอดคล้องกับข้อมูล เมื่อตัดข้อที่ไม่เหมาะสม
โมเดลออกแล้วจึงเหลือหาค่าถามเพียง 55 ข้อ

ผลการวิเคราะห์รายข้อตามโมเดลดั้งเดิม

ตาราง 7 แสดงดัชนีความง่าย (p) และค่าอำนาจจำแนก (r_{11}) ของ
ข้อคำถามทั้ง 60 ข้อ จากผลการสอบของนักเรียนจำนวน 1,502 คน ค่า p ของข้อใด
เป็นอัตราส่วนของจำนวนคนที่ตอบข้อนั้นได้ถูกต้องจำนวนคนที่ตอบข้อนั้น ส่วนค่า r_{11} เป็น
ค่าสัมประสิทธิ์สหสัมพันธ์พ้อยท์ไบซีเรียล ระหว่างข้อที่ i กับคะแนนรวมที่คิดจากทุกข้อ
ในฉบับ ยกเว้นข้อที่ i

ส่วนผลการวิเคราะห์อย่างละเอียดตามแนวการวิเคราะห์รายข้อของหน่วย
บริการทดสอบทางการศึกษา (ETS) ประเทศสหรัฐอเมริกาแสดงไว้ในภาคผนวก จ ใน
การแบ่งคะแนนรวมออกเป็น 5 กลุ่ม ใช้คะแนน 5 ช่วงดังนี้ 11-25, 26-30, 31-34,
35-39 และ 40-57 ซึ่งทำให้จำนวนผู้สอบในแต่ละช่วงใกล้เคียงกันที่สุดเท่าที่จะทำได้
การแสดงผลการนับจำนวนผู้สอบในแต่ละเซลล์ได้ใส่ค่าสัดส่วนคนเลือกตอบข้อเลือกนั้นไว้
ในวงเล็บด้วย (ค่าสัดส่วนนี้คิดมาจากจำนวนคนในเซลล์นั้นหารด้วยจำนวนคนรวมสำหรับ

คอลัมน์นั้น)

ตาราง 7 แสดงดัชนีความง่าย (p) และอำนาจจำแนก (r_{1t}) ของข้อคำถาม 60

ข้อ ในแบบทดสอบวัดผลสัมฤทธิ์ทางคณิตศาสตร์ของนักเรียนชั้นประถมศึกษาปีที่ 6

จำนวน 1,502 คน

ข้อ	p	r_{1t}	ข้อ	p	r_{1t}
1	.9154	.2589	31	.6418	.3559
2	.6145	.3683	32	.6245	.2016
3	.9214	.1807	33	.4075	.1943
4	.4254	.2436	34	.3149	.0664
5	.4348	.2982	35	.4048	.1064
6	.2244	.2093	36	.3682	.2353
7	.8356	.2156	37	.3509	.3144
8	.7943	.2280	38	.2923	.2389
9	.8435	.1609	39	.2071	.2568
10	.5140	.2494	40	.5786	.3169
11	.6984	.2911	41	.1884	.1279
12	.8136	.1865	42	.2929	.2153
13	.2710	.1780	43	.4627	.3237
14	.2437	-.0891	44	.4800	.2658

ตาราง 7 (ต่อ)

ข้อ	p	r _{it}	ข้อ	p	r _{it}
15	.5932	.1367	45	.6372	.4170
16	.5832	.1661	46	.4321	.2194
17	.9401	.1315	47	.8389	.3632
18	.8182	.2036	48	.6911	.3684
19	.7257	.2451	49	.4993	.3766
20	.5186	.2661	50	.6005	.2793
21	.5699	.2350	51	.4061	.2324
22	.2543	.1516	52	.7117	.3869
23	.5972	.3440	53	.4874	.4041
24	.6545	.2572	54	.6944	.2432
25	.2723	.1130	55	.5672	.2774
26	.5313	.3246	56	.5596	.3249
27	.4231	.2648	57	.6200	.2651
28	.3138	.2377	58	.1853	.0648
29	.7282	.3110	59	.7725	.2580
30	.5836	.0736	60	.5023	.0995

ผลจากตาราง 7 แสดงว่าข้อคำถามทั้งหมด 60 ข้อ มีค่า p อยู่ระหว่าง .1853 ถึง .9401 โดยมีค่าเฉลี่ย (p) เป็น .5413 และส่วนเบี่ยงเบนมาตรฐาน (σ_p) เป็น .2004 ข้อที่ค่า p สูงกว่า .90 มีเพียง 3 ข้อ คือข้อ 1, 3 และ 17

ค่าอำนาจจำแนก (r_{1u}) มีค่าอยู่ระหว่าง -.0891 ถึง .4170 ข้อที่มีค่า r_{1u} ต่ำสุด และเป็นค่าลบคือข้อ 14 ข้อที่มีค่า r_{1u} ต่ำกว่า .10 มี 5 ข้อคือ ข้อ 14, 30, 34, 58 และ 60 และข้อที่มี r_{1u} ต่ำกว่า .20 มี 17 ข้อคือข้อ 3, 9, 12, 13, 14, 15, 16, 17, 22, 25, 30, 33, 34, 35, 41, 58, และ 60

เมื่อพิจารณาข้อมูลในภาคผนวก จ โดยมีหลักการว่าข้อคำถามที่ดีควรมีลักษณะดังนี้ เมื่อระดับความสามารถ (คะแนนรวม) สูงขึ้นสัดส่วนคนเลือกตัวเลือกที่ถูกต้องจะสูงขึ้น ส่วนสัดส่วนคนเลือกตัวลวงจะต่ำลง ผลการวิเคราะห์ข้อมูลแสดงว่าข้อที่มีลักษณะไม่สอดคล้องกับหลักดังกล่าวในตัวเลือกเดียวมี 18 ข้อ และ ข้อที่มีลักษณะไม่สอดคล้องกับหลักดังกล่าวในสองตัวเลือก 3 ข้อ ดังปรากฏรายละเอียดในตาราง 8

ตาราง 8 แสดงหมายเลขข้อ (พร้อมทั้งตัวเลือก) ที่มีลักษณะไม่สอดคล้องกับหลักของข้อ
คำถามที่ดี

ตัวเลือกที่	หมายเลขข้อ (ที่มีตัวเลือกไม่สอดคล้องตัวเดียว)
-------------	---

ไม่สอดคล้อง

ก	6, 13, 28, 32, 33, 42, 51, 60
ข	4, 39
ค	41, 43
ง	5, 34, 35, 38, 46, 57

หมายเลขข้อ (ที่มีตัวเลือกไม่สอดคล้องสองตัว)

ก*, ค	14
ก, ข	22
ข*, ค	58

* หมายถึงตัวเลือกที่ถูก

ความสัมพันธ์ระหว่างค่าความยากที่ประมาณจากโมเดลโลจิสติกและโมเดลดั้งเดิม

ความสอดคล้องระหว่างค่าประมาณความยากภายใต้โมเดล 3PL โมเดล 1PL และ โมเดลดั้งเดิม ตรวจสอบได้จากความสัมพันธ์เชิงอันดับที่แสดงในตาราง 9 ชนิดของสหสัมพันธ์ที่ใช้คือสเปียร์แมน โร (Spearman Rho)

ตาราง 9 แสดงค่าสัมประสิทธิ์สหสัมพันธ์แบบอันดับ (สเปียร์แมน โร) ระหว่างค่า
ประมาณความยากจากโมเดล 3PL โมเดล 1PL และโมเดลดั้งเดิมของข้อคำถาม
59 ข้อ ในแบบทดสอบวัดผลสัมฤทธิ์วิชาคณิตศาสตร์

โมเดล	3PL	1PL
1PL	.9784***	
ดั้งเดิม	.9781***	.9998***

*** $p < .0001$

บทที่ 4

สรุปและอภิปรายผล

ผลโดยสรุป

การวิเคราะห์ข้อสอบในครั้งนี้ใช้ข้อมูลที่นักเรียนชายหญิงระดับชั้นประถมศึกษาปีที่ 6 จำนวน 1,502 คน ในเขตบางกะปิ กรุงเทพมหานคร ตอบแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ในสถานการณ์สอบจริงซึ่งเป็นการสอบปลายปีการศึกษา 2529 ของโรงเรียน แบบทดสอบประกอบด้วยข้อคำถามชนิดเลือกตอบ (สี่ตัวเลือก) จำนวนทั้งหมด 60 ข้อซึ่งกลุ่มครูผู้สอนวิชาคณิตศาสตร์เป็นผู้สร้างขึ้น

ผลการตรวจสอบคะแนนรวมจากแบบทดสอบแสดงว่าคะแนนมีมติเดียว

โดยประมาณ และคะแนนมีการกระจายเป็นลักษณะใกล้เคียงสมมาตร

การประมาณค่าพารามิเตอร์ภายใต้โมเดลโลจิสติก ที่มีพารามิเตอร์ 3 ตัว (โมเดล 3PL) โดยใช้โปรแกรม LOGIST ต้องตัดข้อคำถาม ที่ 14 ออกจึงจะให้ผลที่สมบูรณ์ ฉะนั้นผลที่แสดงไว้ในบทที่ 3 จึงเป็นค่าประมาณพารามิเตอร์ของข้อคำถามจำนวน 59 ข้อ ค่าประมาณค่าอำนาจจำแนก (a) มีช่วงจาก .1471 ถึง 2.0000 โดยมีข้อที่มีอำนาจจำแนกต่ำกว่า .2000 อยู่ 2 ข้อ คือ ข้อ 30 และ 60 และข้อที่มีอำนาจจำแนกต่ำกว่า .3000 มี 4 ข้อคือ ข้อ 15, 30, 41, และ 60 ค่าประมาณความยาก (b) มีช่วงจาก -3.7490 ถึง 3.9110 ส่วนค่าประมาณแลซึมเอดล่าง หรือการเดา (c) ที่ต่ำสุดคือ .0418 และสูงสุดคือ .3966 และค่า c ที่โปรแกรมต้องปรับให้คงที่มีถึง 25 ข้อ

การประมาณค่าพารามิเตอร์ภายใต้โมเดลโลจิสติกที่มีพารามิเตอร์ตัวเดียว

(โมเดล 1PL) หรือโมเดลของราซุ โดยใช้โปรแกรม BICAL ต้องตัดข้อ 14 ออกเช่นเดียวกัน เนื่องจากไม่เหมาะสมกับโมเดลอย่างชัดเจนหลังจากตัดข้อ 14 ออกแล้วจึงวิเคราะห์ 59 ข้อที่เหลือ พบว่าค่าประมาณความยากมีช่วงจาก -2.68 ถึง 1.90

และในการวิเคราะห์ครั้งหลังนี้พบข้อที่ไม่เหมาะสมกับโมเดลอยู่อีก 4 ข้อ คือ ข้อ 30, 34, 45, และ 60 เมื่อพิจารณาได้งแสดงลักษณะข้อคำถามในภาคผนวก ก ประกอบกับ 08.1 และ 80.3- การวิเคราะห์ผลการทดสอบพบว่าข้อ 30, 34, และ 60 มีค่าที่พลัดจากค่า p แบบรายชื่อ ส่วนข้อ 45 มีค่าที่ผิดเพี้ยนมาก ซึ่งเมื่อย้อนกลับไปพิจารณาค่าอำนาจจำแนกของข้อเหล่านี้ในตาราง 6 แล้ว จึงพบว่าข้อ 30, 34, และ 60 มีค่าอำนาจจำแนกต่ำมาก (.33 หรือ .34) ส่วน ข้อ 45 มีค่าอำนาจจำแนกสูงมาก (1.71) ค่าอำนาจจำแนกของข้อที่ควรจะเป็นตาม โมเดลคือ 1.00 ดังนั้นข้อคำถามทั้งสี่ข้อนี้จึงมีพฤติกรรมไม่สอดคล้องกับโมเดลกล่าว โดยสรุปข้อคำถามที่มีอำนาจจำแนกสูงมาก และต่ำมาจาก 1 จะไม่เหมาะสมกับโมเดล

ในโมเดลดั้งเดิมผลการประมาณค่าความง่าย (p) พบว่ามีค่าระหว่าง .1853 ถึง .9401 ข้อที่มีค่า p อยู่นอกช่วง (.10 - .90) มี 3 ข้อคือ ข้อ 1, 3 และ 17 ค่าประมาณอำนาจจำแนก (r_{11}) มีค่าระหว่าง -.0891 ถึง .4170 โดย ข้อ 14 มีค่าอำนาจจำแนกต่ำสุดและเป็นค่าลบ ข้อที่มีค่าอำนาจจำแนกต่ำกว่า .10 ($r_{11} < .10$) มี 5 ข้อ คือ ข้อ 14, 30, 34, 58, และ 60 ส่วนข้อที่มีค่าอำนาจจำแนกต่ำกว่า .20 ($r_{11} < .20$) มี 17 ข้อ คือ ข้อ 3, 9, 12, 13, 14, 15, 16, 17, 22, 25, 30, 33, 34, 35, 41, 58, และ 60

ผลการวิเคราะห์ข้อส่างละเอียด ดูในภาคผนวก จ แนวการวิเคราะห์นี้มี หลักว่า ข้อคำถามควรมีลักษณะดังนี้ เมื่อคะแนนรวมสูงขึ้นสัดส่วนคนเลือกตัวเลือกที่ถูก จะสูงขึ้น ส่วนสัดส่วนคนเลือกตัวลวงจะต่ำลง ผลนี้แสดงว่าข้อที่ผลการตอบตัวเลือกไม่ เป็นไปตามหลักการมีหลายข้อ โดยแบ่งออกได้เป็น 2 ประเภท คือ ข้อที่มีปัญหาในตัว เลือกเดียว และข้อที่มีปัญหาใน 2 ตัวเลือก สำหรับข้อที่มีปัญหาใน 2 ตัวเลือกคือข้อ 14, 22, และ 58 (ดูตาราง 8) โดยเฉพาะข้อ 14 และ 58 มีปัญหาในตัวเลือก ที่ถูกต้อง ส่วนข้อที่มีปัญหาในตัวเลือกเดียวนั้นมี 18 ข้อ (ดูรายละเอียดในตาราง 8)

อภิปรายผล

การวิเคราะห์ข้อคำถามในแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนครั้งนี้ให้
ผลการตอบข้อคำถามของนักเรียน 1,502 คน ในสถานการณ์ที่เป็นการสอบจริงใน
โรงเรียน ดังนั้นคะแนนที่ใช้นี้จึงสะท้อนความพยายามสูงสุดของนักเรียน

ความสัมพันธ์ระหว่างค่าประมาณความยากในโมเดลทั้งสาม

ทั้งโปรแกรม LOGIST และโปรแกรม BICAL ให้วิธีการไล่ลำดับสูงสุด
ตามแนวของเบิร์นบอม (Birnbbaum, 1968) ในการประมาณค่าพารามิเตอร์ของข้อ และ
ของผู้สอบพร้อมกันดังได้กล่าวแล้วในบทที่ 1 แต่ในโปรแกรมทั้งสองมีวิธีการปรับมาตรา
ความสามารถที่แตกต่างกัน โปรแกรม LOGIST แปลงเชิงเส้นตรงให้มาตราของ
ค่าประมาณความสามารถมีค่าเฉลี่ยเป็นศูนย์ (0) และส่วนเบี่ยงเบนมาตรฐานเป็น 1
(Wood, et al., 1967; and Wingersky et al., 1982) ส่วนโปรแกรม
BICAL ปรับตำแหน่งของความสามารถโดยตั้งให้ค่าเฉลี่ยของค่าประมาณความยาก
เป็นศูนย์ (0) (Wright and Douglas, 1977) ดังนั้นค่าพารามิเตอร์ความยาก
ที่ประมาณจากโปรแกรมทั้งสองจึงแตกต่างกันและแม้ว่าจะใช้โปรแกรม LOGIST
ประมาณค่าความยากภายใต้โมเดล 1PL ด้วยก็ตาม ค่าประมาณพารามิเตอร์ดังกล่าว
ก็จะแตกต่างจากที่ได้จากโมเดล 3PL เมื่อพิจารณาอันดับของข้อตามค่าความยากที่
ประมาณได้ภายใต้โมเดล 3PL (โดยใช้โปรแกรม LOGIST) และโมเดล 1PL
(ใช้โปรแกรม BICAL) ในตารางในภาคผนวก จ พบว่าภายใน 10 อันดับแรก
(เรียงลำดับจากค่าความยากต่ำสุดไปหาสูงสุด) เป็นข้อคำถามข้อเดียวกันแต่อันดับ
สลับกันไปบ้างสำหรับอันดับของทั้ง 59 ข้อ คำนวณค่าสหสัมพันธ์แบบอันดับตาม
สูตรสเปียร์แมน r_s พบว่าสัมประสิทธิ์สหสัมพันธ์มีค่า 0.9784 ซึ่งเป็นค่าที่มีนัยสำคัญ
ทางสถิติที่ .0001 สำหรับการทดสอบด้านเดียว นี่เป็นหลักฐานที่แสดงว่าค่าประมาณ
ความยากภายใต้โมเดลทั้งสองมีอันดับใกล้เคียงกันมาก

ส่วนอันดับของข้อความทั้ง 59 ข้อเรียงตามค่าความยากที่คำนวณภายใต้โมเดลดั้งเดิม (ค่า p สูงขึ้น ข้อความยิ่งง่ายขึ้น ค่า p ต่ำลงข้อความยิ่งยากขึ้น) แสดงไว้ในตารางในภาคผนวก จ เช่นเดียวกัน (เพื่อความสะดวกในการเปรียบเทียบ จึงตัดข้อ 14 ออกด้วย) ค่าสัมประสิทธิ์สหสัมพันธ์แบบอันดับตามสูตรสเปียร์แมน ระหว่างอันดับที่จัดตามโมเดล 3PL กับอันดับที่จัดตามโมเดลดั้งเดิมมีค่า .9781 และค่าสัมประสิทธิ์สหสัมพันธ์ชนิดเดียวกัน (สเปียร์แมน ρ) ระหว่างอันดับที่จัดตามโมเดล 1PL กับอันดับที่จัดตามโมเดลดั้งเดิมมีค่า .9998 ค่าสัมประสิทธิ์สหสัมพันธ์ทั้งสองมีนัยสำคัญทางสถิติที่ระดับ .0001 สำหรับการทดสอบด้านเดียว

ผลการคำนวณค่าความสัมพันธ์เชิงอันดับแสดงให้เห็นว่าความยากของข้อตามโมเดลดั้งเดิมมีความสัมพันธ์ทางบวกกับโมเดล 1PL สูงมาก การที่ความยากสอดคล้องกันมากเช่นนี้เป็นสาเหตุหนึ่งที่มีผู้ไม่เห็นด้วยกับการใช้โมเดล 1PL (Kline, 1986) อันดับความยากของข้อภายใต้โมเดลดั้งเดิมมีความสัมพันธ์ทางบวกกับอันดับความยากภายใต้โมเดล 3PL ต่ำกว่าที่สัมพันธ์กับอันดับความยากภายใต้โมเดล 1PL เนื่องจากในกรณีทั่วไปที่อำนาจจำแนก (a) ไม่เท่ากันนั้น การจัดอันดับตามค่า p กับตามค่าของ b จะไม่เหมือนกันแต่ถ้าทุกข้อมีอำนาจจำแนกเท่ากัน (a เท่ากัน) ค่า b และ p มีความสัมพันธ์กันในลักษณะที่ว่าเมื่อ p เพิ่มขึ้น (ง่ายขึ้น) ค่า b จะลดลง (ง่ายขึ้น) (Lord, 1980, p.34; พงษ์จิต อินทสุวรรณ, 2528, หน้า 47-48) ดังนั้นอันดับของข้อความทั้ง 59 ข้อที่จัดตามค่าความยากภายใต้โมเดล 1PL กับที่จัดตามค่าความยากภายใต้โมเดลดั้งเดิมจึงตรงกันแทบทุกข้อ มีเพียง 6 ข้อเท่านั้นที่ไม่ตรง ข้อเหล่านี้แตกต่างกันอย่างมีระบบคือ มีอันดับติดกันแต่สลับกันเป็นคู่ ๆ ดังต่อไปนี้ ลำดับของข้อ 16 และ 30 สลับกัน (อันดับที่ 26 และ 27) ข้อ 34 และ 28 สลับกัน (อันดับที่ 49 และ 50) และสุดท้ายข้อ 42 และ 38 สลับกัน (อันดับที่ 51 และ 52) หลักฐานที่ปรากฏ ที่สามารถใช้เป็นคำอธิบายของ

การสลับข้อกันนี้มี 2 ประเด็น ประเด็นแรก ค่าประมาณความยากภายใต้โมเดล 1PL จากโปรแกรม BICAL ที่พิมพ์ไว้มีค่าเท่ากัน (สำหรับคู่แรกและคู่สุดท้าย) และอีกประเด็นหนึ่งคือ (ในคู่แรกและคู่ที่สอง) มีข้อที่ไม่เหมาะสมกับโมเดล 1PL (ดูตาราง 6) ข้อที่ไม่เหมาะสมคือ ข้อ 30 และ 34

พิจารณาความสอดคล้องของอันดับของอำนาจจำแนกที่เป็นผลจากการวิเคราะห์ภายใต้โมเดล 3PL (a เป็นอำนาจจำแนก) กับผลจากการวิเคราะห์ภายใต้โมเดลดั้งเดิม (r_{1u} เป็นอำนาจจำแนก) พบว่าไม่มีแนวโน้มที่จะสอดคล้องกัน ซึ่งโดยหลักการแสดงให้เห็นอยู่แล้วว่ามีโอกาสน้อยมากที่จะมีความสัมพันธ์กัน โดยเฉพาะอย่างยิ่งเมื่อ $c \neq 0$ (Lord, 1980, p.33; ผจจจิต อินทสุวรรณ, 2528. หน้า 46)

ปัญหาของข้อคำถามบางข้อ

ก. พิจารณาผลจากโมเดลทั้งสาม

จากข้อคำถามทั้งหมด 60 ข้อ ในแบบทดสอบที่วิเคราะห์ในครั้งนี้มีข้อคำถามอยู่ 3 ข้อที่ทุกวิธีระบุว่ามีปัญหา ข้อแรกคือ ข้อ 14 เป็นข้อที่ผลการตอบมีปัญหาเด่นชัดที่สุดนั่นคือทำให้การวิเคราะห์โดยใช้โปรแกรม LOGIST ไม่สมบูรณ์ต้องตัดออกจากการวิเคราะห์ เป็นข้อที่โปรแกรม BICAL (ดูตาราง 5) ให้ค่าสถิติสำหรับทดสอบความเหมาะสมกับโมเดล 1PL สูงเด่นชัดที่สุด ($t=4.63$, $b=1.48$, Disc index = $-.16$) และเป็นข้อที่ผลการวิเคราะห์ตามโมเดลดั้งเดิมให้ค่าอำนาจจำแนก (r_{1u}) ต่ำสุด และเป็นค่าลบ ($r_{1u} = -.0891$) ค่า p เท่ากับ $.2437$ นอกจากนี้ยังเป็นข้อที่มีตัวเลือกทั้ง 2 ตัว (ตัวเลือก 1 และ 3) มีพฤติกรรมที่ไม่สอดคล้องกับหลักการวิเคราะห์ตามแบบของ ETS ตัวเลือก 1 เป็นตัวถูกต้องตามเฉลย(ที่ตั้งใจให้คลาดเคลื่อน) แต่เมื่อเปรียบเทียบกับระหว่างนักเรียน 5 กลุ่ม ที่มีความสามารถจากต่ำไปสูง แล้วพบว่าในกลุ่มนักเรียนที่มีความสามารถสูงสุด สัดส่วนการเลือกตอบตัวถูกนี้ต่ำกว่าในกลุ่มอื่น ๆ สำหรับตัวเลือก 3 สัดส่วนการเลือกตอบกลับเพิ่มสูงขึ้นตามระดับความสามารถของ

กลุ่มนักเรียน นักเรียนเลือกตอบตัวเลือก 3 กันมากที่สุด และสูงกว่าการเลือกตอบตัวถูกอย่างเด่นชัด

ข้อคำถามข้อ 14 เป็นข้อคำถามในเนื้อเรื่องของรูปสมมาตร โจทย์ถามให้นักเรียนระบุรูปที่มีแกนสมมาตรเป็นจำนวนเท่ากับรูปสี่เหลี่ยมผืนผ้า ซึ่งคำตอบที่ถูกที่สุดเป็นข้อเลือก 3 ส่วนข้อเลือก 1 นั้น มีแกนสมมาตรเป็นจำนวนมากกว่ารูปสี่เหลี่ยมผืนผ้า (แต่เมื่อพิจารณาผิวเผินอาจคิดว่าเท่ากันได้) ปัญหาข้อนี้จึงเป็นการเฉลยคำตอบไม่ถูกต้อง หลักฐานนี้เป็นเครื่องแสดงถึงประโยชน์ส่วนหนึ่งของ การวิเคราะห์รายข้อ ซึ่งจะช่วยให้ผู้สร้างข้อสอบได้ตรวจสอบซ้ำในปัญหาเกี่ยวกับ ความคลาดเคลื่อนในการเฉลยคำตอบ

การอภิปรายผลจากโปรแกรม BICAL สำหรับทุกข้อ (ยกเว้นข้อ 14)

ใช้ข้อมูลในตาราง 6

ข้อคำถามข้อ 30 มีอำนาจจำแนกต่ำสุด ($a=.1471$) ค่าความยาก (b) เท่ากับ $-.0381$ และการเดา (c) เป็น $.1646$ ภายใต้โมเดล 3PL เป็นข้อที่ไม่เหมาะสมกับโมเดล 1PL ในลักษณะที่ค่าอำนาจจำแนกต่ำกว่า 1 มากไป ($t=2.90$, $b=-.11$, Disc index $=.33$) และมีอำนาจจำแนกภายใต้โมเดลดั้งเดิมต่ำมาก ($r_{11}=.0736$) ค่า p เท่ากับ $.5836$ เนื้อหาในข้อ 30 เป็นบทประยุกต์ เป็นคำถามให้แก้ปัญหาเกี่ยวกับการซื้อของที่ต้องใช้การคูณและหาร ผลจากตารางในภาคผนวก จ แสดงว่าถึงแม้ว่านักเรียนจะเลือกตอบข้อเลือก 2 กันมาก แต่คะแนนของนักเรียนที่เลือกตัวถูก (ตัวเลือก 2) สูงกว่าค่าเฉลี่ยของนักเรียนที่เลือกตอบตัวลงทั้งสามเพียงเล็กน้อย (ตัวเลือกทั้งสี่ ให้เทอมเดียวกันสลับที่กันในการคูณและหาร) สรุปได้ว่าข้อนี้เป็นข้อยากปานกลางแต่ไม่มีอำนาจจำแนก นักเรียนในกลุ่มเก่งและอ่อนมีสัดส่วนการเลือกตอบใกล้เคียงกัน

ข้อ 60 มีอำนาจจำแนกต่ำถัดจาก ข้อ 30 ($a=.1716$) ค่าความยาก (b) เท่ากับ 1.3370 และการเดา (c) เป็น .1646 ภายใต้อาณาเขต 3PL เป็นข้อที่ไม่เหมาะสมกับโมเดล 1PL ในลักษณะเดียวกันกับข้อ 30 อย่างเด่นชัด ($t=3.03$, $b=.25$, Disc index = .34) ค่าอำนาจจำแนกภายใต้อาณาเขตเดิมมีค่าต่ำมาก ($r_{1t}=0.0995$ และ ค่า p เท่ากับ .5023 นอกจากนี้สัดส่วนการเลือกตัวเลือก 1 ยังมีทิศทางกลับกับหลักที่ควรจะเป็นอีกด้วย (ดูภาคผนวก จ)

จากการเฉลยคำตอบที่กำหนดให้คลาดเคลื่อนไว้คือ คำตอบที่ถูกต้องเป็นตัวเลือก 1 ไม่ใช่ตัวเลือก 4 ดังที่เฉลยไว้ ดังนั้นแนวโน้มของสัดส่วนการตอบตัวเลือก 1 จึงเป็นสิ่งที่ถูกต้อง คือมีแนวโน้มเพิ่มขึ้นตามระดับความสามารถ อย่างไรก็ตามนักเรียนเลือกตอบตัวเลือก 4 มากที่สุด ในทุกระดับความสามารถ และสัดส่วนการเลือกตอบตัวเลือกนี้ยังมีทิศทางเพิ่มขึ้นตามระดับความสามารถอีกด้วย (ดูภาคผนวก จ) หลักฐานการตอบของนักเรียนนี้แสดงว่านักเรียนจำนวนหนึ่งยังมีปัญหาเกี่ยวกับทศนิยมหรือมีจะนั้นก็ปัญหาเรื่องการเทียบมาตราจากเมตรเป็นกิโลเมตร (ตัวเลือก 1 และ 4 มีค่า 3.80 และ 38.00 กิโลเมตร ตามลำดับ)

ข. พิจารณาผลจากโมเดล 3 PL

เมื่อพิจารณาผลจากวิเคราะห์จากโปรแกรม LOGIST พบว่าข้อที่มีปัญหามากที่สุดคือ ข้อ 14 ซึ่งทำให้การวิเคราะห์ไม่สมบูรณ์ ส่วนอีก 2 ข้อคือ ข้อ 30 และ 60 มีค่าอำนาจจำแนก (a) ต่ำกว่า .20 ข้อคำถามทั้ง 3 ข้อ นี้ได้อภิปรายผ่านมาแล้ว

ค. พิจารณาผลจากโมเดล 1PL

สำหรับผลการวิเคราะห์จากโปรแกรม BICAL พบว่ามีข้อคำถามที่ไม่เหมาะสมกับโมเดล 1PL รุนแรงที่สุด 5 ข้อคือข้อ 14, 30, 34, 45, และ 60 ซึ่งข้อ 14, 30, และ 60 นั้น ได้อภิปรายในรายละเอียดมาแล้ว ต่อไปนี้จะพิจารณาถึงปัญหาของข้อ 34 และ 45

ข้อ 34 มีค่าอำนาจจำแนกตามผลจาก LOGIST เป็น 2.00 (ดูตาราง 4) ซึ่งสูงกว่าข้ออื่นใดในแบบทดสอบ และเป็นค่าสูงสุดของ a ที่โปรแกรมกำหนดไว้ (ค่าที่ประมาณสูงกว่า 2.00 แต่โปรแกรมบังคับค่าไว้ที่ 2.00 แสดงว่าค่าที่ได้ไม่ดี) ค่าความยาก (b) เท่ากับ 2.6232 แสดงว่าเป็นข้อที่ยากมากข้อหนึ่งถึงแม้จะมีอำนาจจำแนกสูง (ดูโค้งในภาคผนวก ข) ผลจากโปรแกรม BICAL แสดงว่าข้อนี้ไม่เหมาะกับโมเดล 1PL เนื่องจากค่าอำนาจจำแนกต่ำกว่าที่คาดไว้มาก ($t=2.91$, $b=1.12$, $\text{Disc index} = .33$ ภายหลังโมเดลดั้งเดิมมีอำนาจจำแนก (r_{1u}) ต่ำมากคือมีค่าเท่ากับ .0664 และค่า p เท่ากับ .3149 พิจารณาผลในภาคผนวก จ พบว่าค่าเฉลี่ยของนักเรียนที่ตอบถูกมีค่าสูงกว่าค่าเฉลี่ยของนักเรียนที่ตอบตัวลวงเพียงเล็กน้อย

ผลจากการวิเคราะห์ทั้งสามแนวทางสรุปได้ตรงกันว่าข้อนี้ค่อนข้างยาก แต่ขัดแย้งกันในเรื่องอำนาจจำแนก นั่นคือโปรแกรม LOGIST แสดงค่าอำนาจจำแนกสูงมาก ส่วนผลจากโปรแกรม BICAL และจากโมเดลดั้งเดิมแสดงค่าต่ำมาก เหตุการณ์นี้สามารถอธิบายได้โดยการสังเกตค่าประมาณการเดา (c) (หรือ แอซิมโทตล่าง) จากตาราง 4 ค่า c สำหรับข้อนี้เท่ากับ .3063 ซึ่งสูงกว่าโอกาสในการเดาอย่างสุ่ม (.25) ค่านี้เป็นค่าที่สูงมาก ส่วนโมเดล 1PL นั้นสมมติให้ c มีค่าเป็น 0 (ไม่มีการเดา) ดังนั้นค่าประมาณอำนาจจำแนกจึงมีทิศทางที่ขัดแย้งกันดังนั้นจึงสรุปได้ว่าการที่ข้อ 34 ไม่เหมาะกับโมเดล 1PL เนื่องจากการเดาที่มีค่าสูงมากนั่นเอง ผลที่พบนี้สอดคล้องกับการศึกษาของ Reckase (1979) ที่พบว่าการเดาเป็นส่วนประกอบสำคัญของสถิติที่ใช้ทดสอบความเหมาะสมในโมเดล 1PL โมเดลดั้งเดิมซึ่งให้คะแนน 0-1 ที่ให้ในงานวิจัยนี้ก็ไม่ได้คำนึงถึงการเดาเช่นกัน อีกประการหนึ่งอำนาจจำแนกตามโมเดลดั้งเดิมมีผลกระทบจากความสามารถของกลุ่มผู้สอบด้วย กล่าวโดยสรุปโมเดลที่บรรยายข้อนี้ได้ดีกว่าคือโมเดล 3PL

สำหรับข้อ 45 ผลจากโปรแกรม LOGIST ให้ค่าอำนาจจำแนก (a) เท่ากับ 1.0955 ความยาก (b) เป็น -.2740 และการเดา (c) เป็น .1371 ผลจากโปรแกรม BICAL แสดงว่าข้อนี้ไม่เหมาะกับโมเดล 1PL ประมาณค่าความยากได้เท่ากับ -.36 และค่าอำนาจจำแนกสูงถึง 1.71 (ควรเป็น 1) ส่วนผลของการวิเคราะห์ตามโมเดลดั้งเดิมพบว่าค่าความง่าย (p) เท่ากับ .6372 และอำนาจจำแนก (r_{1c}) เป็น .4170 นอกจากนี้ผลในภาคผนวก จ ก็เป็นไปตามแบบแผนที่เหมาะสม จากข้อมูลทั้งหมดนี้แสดงว่าข้อนี้ไม่เหมาะกับโมเดล 1PL เนื่องจากอำนาจจำแนกที่สูงมาก

ง. พิจารณาผลจากโมเดลดั้งเดิม

เมื่อพิจารณาผลจากการวิเคราะห์ภายใต้โมเดลดั้งเดิม ซึ่งเป็นการคำนวณค่า p และ r_{1c} รวมทั้งการตรวจสอบแบบแผนการตอบข้อเลือกตามลำดับความสามารถด้วยปรากฏว่าไม่พบข้อที่เป็นปัญหามากสำหรับผลการคำนวณค่า p ส่วนผลการคำนวณค่า r_{1c} พบว่ามีข้อที่มีอำนาจจำแนกต่ำมาก ($r_{1c} < .10$) จำนวน 5 ข้อคือ ข้อ 15, 30, 34, 58, และ 60 นอกจากนี้ข้อ 14, 22, และ 58 ยังมีปัญหาเกี่ยวกับลักษณะการตอบข้อเลือกต่าง ๆ ที่ไม่เป็นไปตามหลักการถึง 2 ข้อเลือกด้วย ข้อ 14, 30, 34, และ 60 นั้น ได้อภิปรายมาแล้ว ต่อไปนี้จะอภิปรายถึงข้อ 22 และ 58

ข้อ 22 มีค่า a เท่ากับ 1.0847 ค่า b เท่ากับ 2.1845 และค่า c .2065 ภายใต้โมเดล 3PL เป็นข้อที่เหมาะสมกับโมเดล 1PL โดยมีค่า b เท่ากับ 1.44 ส่วนผลการวิเคราะห์ภายใต้โมเดลดั้งเดิมปรากฏค่า p เท่ากับ .2543 และค่า r_{1c} เป็น .1516 ผลเหล่านี้แสดงว่า เมื่อวิเคราะห์ภายใต้โมเดลโลจิสติกไม่พบปัญหาผิดปกติ แต่เป็นปัญหาในการใช้โมเดลดั้งเดิม เนื่องจากพารามิเตอร์ภายใต้โมเดลดั้งเดิมขึ้นอยู่กับกลุ่มผู้สอน

โจทย์ข้อ 22 เกี่ยวกับปัญหาการแบ่งทองซึ่งจะต้องให้ทหาร แต่นักเรียนยังไม่เข้าใจเรื่องนี้ข้างละเอียดพอ และตัวเลือก 1 และ 2 ให้คำว่า "แบ่ง" เช่นเดียวกับตัวเลือก 3 ซึ่งเป็นตัวถูก ด้วยเหตุนี้จึงมีนักเรียนกลุ่มกลาง 3 กลุ่มเลือกตอบตัวลงทั้งสองนี้กันมาก ทำให้สัดส่วนการเลือกตอบสูงขึ้น ผู้ที่เลือกตอบตัวลงทั้งสองนี้มีจำนวนสูงกว่าผู้ที่เลือกตอบตัวถูก นอกจากนี้คะแนนเฉลี่ยของกลุ่มผู้สอบที่เลือกตัวเลือกที่ถูกยังสูงกว่าคะแนนเฉลี่ยของกลุ่มผู้ที่เลือกตัวลงทั้งสองดังกล่าวไม่มากนัก เป็นเหตุให้ค่าอำนาจจำแนก (r_{1c}) ที่คำนวณได้มีค่าต่ำ

กล่าวโดยสรุป โมเดลดั้งเดิมซึ่งค่าประมาณพารามิเตอร์ขึ้นอยู่กับกลุ่มผู้สอบ และในโจทย์ข้อนี้ผู้สอบกลุ่มนี้ไม่สามารถจำแนกแยกแยะตัวถูกออกจากตัวลงได้ ดังนั้นข้อคำถามข้อนี้จึงไม่ใช่ข้อที่ดีเมื่อวิเคราะห์ด้วยโมเดลดั้งเดิม ส่วนโมเดลโลจิสติกทั้งโมเดล 3PL และโมเดล 1PL ให้ผลที่แสดงว่าข้อนี้ให้ได้อย่างไรก็ดีการวิเคราะห์รายข้อตามแบบโมเดลดั้งเดิมทำให้ได้ข้อมูลเฉพาะกับกลุ่มผู้สอบอันจะนำไปสู่การปรับปรุงข้อคำถาม ตัวเลือก และเพื่อความเข้าใจกลุ่มผู้สอบด้วย

ข้อ 58 มีค่าอำนาจจำแนก (a) .9005 ค่าความยาก (b) 2.9089 และการเดา (c) .1576 ภายใต้มอดล 3PL ผลการวิเคราะห์ภายใต้มอดล 1PL แสดงว่าเหมาะสมกับโมเดลโดยมีค่าความยาก (b) เป็น 1.90 ผลที่กล่าวมาแล้วแสดงว่าข้อนี้ไม่มีปัญหาจากการวิเคราะห์ภายใต้มอดลโลจิสติกทั้งสอง ส่วนผลการวิเคราะห์ภายใต้มอดลดั้งเดิม มีค่า p เป็น .1853 และค่า r_{1c} เป็น .0648 แสดงว่าเป็นข้อที่ยากและมีอำนาจจำแนกต่ำมาก นอกจากนี้ผลในภาคผนวก จ ยังแสดงว่าสัดส่วนการตอบตัวเลือกไม่เป็นไปตามหลักการถึง 2 ตัวเลือก คือตัวเลือก 2 และ 3 ซึ่งตัวเลือก 2 เป็นตัวเลือกที่ถูกตามเฉลย

ข้อ 58 เป็นโจทย์เกี่ยวกับการคำนวณระยะทางจากแผนผัง โดยให้มาตราส่วนตามที่กำหนด จากนักเรียนทั้งหมด 1,502 คน และทำข้อนี้ 1,441 คน

มีนักเรียนเลือกตอบตัวเลือก 3 มากที่สุดถึง 715 คน (คิดเป็น 49.6%) ส่วนตัวเลือก 2 ที่เป็นตัวถูกนั้นมีนักเรียนเพียง 269 คน (18.7%) เท่านั้นที่เลือกตอบ การที่นักเรียนเลือกตอบตัวเลือก 3 กันมากนั้น สามารถวิเคราะห์สาเหตุได้ว่านักเรียนอ่านคำถามไม่ละเอียดเพียงพอ หรือมีฉะนั้นก็มีเวลาในการทำข้อนี้น้อยเนื่องจากเป็นข้อท้าย ๆ ของฉบับ

ผลการวิเคราะห์ข้อ 58 นี้ เป็นหลักฐานอีกข้อหนึ่งที่แสดงให้เห็นว่าโมเดลโลจิสติกไม่ได้ชี้ให้เห็นว่ามีปัญหาแต่ประการใด แต่การวิเคราะห์ตามโมเดลดั้งเดิมซึ่งมีการวิเคราะห์รูปแบบการตอบตัวเลือกแต่ละตัวด้วยนั้นชี้ให้เห็นถึงปัญหาของนักเรียนในการเลือกตอบตัวเลือกต่าง ๆ ซึ่งเป็นปัญหาเฉพาะกลุ่มตัวอย่างที่ทดสอบ

สรุปการอภิปรายผล

สรุปข้อคำถามทั้ง 60 ข้อ ในแบบทดสอบวัดผลสัมฤทธิ์วิชาคณิตศาสตร์ สำหรับชั้นประถมศึกษาปีที่ 6 ที่นักเรียนกลุ่มนี้ตอบมาแล้ว และนำมาวิเคราะห์รหัสข้อด้วยโมเดล 3PL โมเดล 1PL และโมเดลดั้งเดิมผลจากการวิเคราะห์สามารถสรุปได้ดังนี้

1. มีข้อคำถาม 3 ข้อคือ ข้อ 14, 30 และ 60 ที่ทั้ง 3 โมเดลให้ผลที่สอดคล้องกันว่าเป็นข้อที่หยังไม่ควรนำไปใช้ โดยที่ข้อ 14 มีการเฉลยผิด ข้อ 30 ไม่มีอำนาจจำแนก นักเรียนมีปัญหาในเรื่องบัญญัติไตรยางค์ และข้อ 60 มีปัญหาในเรื่องการเฉลยผิด และ นักเรียนมีปัญหาในเรื่องทศนิยมด้วย

2. ผลจากการวิเคราะห์โดยใช้โปรแกรม LOGIST (ภายใต้โมเดล 3PL) แสดงว่ามีข้อคำถามที่ไม่ควรนำไปใช้ข้อ 3 ข้อ ซึ่งเป็นข้อเดียวกันกับที่กล่าวแล้วในข้อ 1.

3. ผลจากการวิเคราะห์ภายใต้โมเดล 1PL โดยใช้โปรแกรม BICAL แสดงว่ามีข้อคำถามที่ไม่เหมาะกับโมเดลอยู่ 5 ข้อคือ ข้อ 14, 30, 34, 45, และ 60 ในจำนวน 5 ข้อนี้ ข้อ 14, 30, และ 60 ได้กล่าวถึงมาแล้วในข้อ 1. ส่วน

ข้อ 34 และ 45 นั้น ไม่เหมาะกับโมเดล 1PL เนื่องจากค่าประมาณพารามิเตอร์ การเดาหรือแอซิมโทติกล่าง (c) สูงมาก (ดูผลในตาราง 4) และค่าอำนาจจำแนก สูงเกินไป (ดูตาราง 6) ตามลำดับ ซึ่งโมเดล 1PL มีข้อตกลงเบื้องต้นให้ค่าพารามิเตอร์ ทั้งสองนี้คงที่ นั่นคือ $c=0$ และ a คงที่ ซึ่งในโปรแกรม BICAL กำหนดให้ $a=1$

4. ผลจากการวิเคราะห์ตามแบบโมเดลดั้งเดิม มีข้อคำถามที่ยังไม่ควรให้อยู่ทั้งหมด 6 ข้อคือ ข้อ 14, 22, 30, 34, 58, และ 60 ในจำนวน 6 ข้อนี้ มีข้อที่ไม่เข้ากับข้อที่ได้กล่าวมาแล้วเพียง 2 ข้อคือ ข้อ 22 และ 58 ทั้งสองข้อนี้มีค่าอำนาจจำแนก (r_{1c}) ต่ำมาก และประการสำคัญคือมีรูปแบบการตอบตัวเลือกที่ไม่เป็นไปตามทิศทางที่ถูกต้อง ซึ่งเป็นปัญหาการเลือกใช้ค่าในตัวเลือกหรือปัญหาความไม่เข้าใจเรื่องที่สอบถามอย่างชัดเจน นอกจากนี้ยังอาจมีปัญหาเกี่ยวกับความรอบคอบในการอ่านโจทย์ของผู้สอบด้วย

ผลทั้งหมดนี้แสดงให้เห็นถึงประโยชน์ของการวิเคราะห์รายข้อ โดยใช้โมเดลโลจิสติกและโมเดลดั้งเดิม โมเดลโลจิสติกที่มีพารามิเตอร์ 3 ตัว (โมเดล 3PL) ปฏิเสธข้อคำถามที่วิเคราะห์เป็นจำนวนน้อยกว่าโมเดลของราช (โมเดล 1PL) เนื่องจากพารามิเตอร์ของข้อทั้ง 3 ตัวแปรค่าได้ ส่วนโมเดลของราชกำหนดข้อตกลงเบื้องต้นให้พารามิเตอร์อำนาจจำแนกคงที่ (BICAL กำหนดให้เป็น 1) และการเดาเป็น 0 นั่นคือผู้สอบไม่สามารถตอบถูกต้องด้วยการเดา ดังนั้นจึงปฏิเสธข้อคำถามที่วิเคราะห์หลายข้อที่ไม่เป็นไปตามข้อตกลงเบื้องต้นดังกล่าว ส่วนโมเดลดั้งเดิมนั้นค่าประมาณพารามิเตอร์ของข้อคือความง่าย (p) และอำนาจจำแนก (r_{1c}) ขึ้นอยู่กับกลุ่มผู้สอบ ผลการวิเคราะห์จึงที่เฉพาะกลุ่มผู้สอบ ซึ่งให้ประโยชน์ในด้านของความเข้าใจกลุ่มผู้สอบด้วยนอกเหนือจากความเข้าใจข้อคำถาม นอกจากนี้การวิเคราะห์รูปแบบการตอบตัวเลือกให้ข้อมูลที่ เป็นประโยชน์สำหรับการพิจารณาตัวเลือกที่เป็นตัวลวงและตัวถูกที่เหมาะสม เพื่อการพิจารณาปรับปรุงตัวเลือกให้ดีขึ้นและทำการวิเคราะห์รายข้อซ้ำอีกเพื่อให้ได้แบบทดสอบ

ที่มีคุณภาพดี อันเป็นจุดมุ่งหมายสำคัญของการพัฒนาแบบทดสอบ

ข้อเสนอแนะสำหรับการปฏิบัติ

ผลจากการศึกษาในครั้งนี้จึงเป็นการศึกษาจากข้อมูลที่ได้จากสภาพการณ์
สอบจริงของโรงเรียน เป็นหลักฐานที่ชัดเจนถึงความเหมาะสมของการประยุกต์โมเดล
โลจิสติก และโมเดลดั้งเดิม ดังต่อไปนี้

1. โมเดลโลจิสติกที่มีพารามิเตอร์ 3 ตัว (โมเดล 3PL) สอดคล้อง
กับธรรมชาติของข้อสอบแบบเลือกตอบวัดผลสัมฤทธิ์ทางการเรียน ข้อคำถามที่โมเดล 3PL
ระบุว่าให้ไม่ได้นั้นมีจำนวนน้อย และเป็นปัญหาสำคัญจริง ๆ ต้องแก้ไข อย่างไรก็ตามผู้วิจัย
ต้องสังเกตผลการประมาณค่าพารามิเตอร์ของโปรแกรม LOGIST สำหรับ a และ c ซึ่งอาจ
ได้ค่าที่ไม่ดี

2. โมเดลโลจิสติกที่มีพารามิเตอร์ 1 ตัว (โมเดล 1PL หรือโมเดล
ของราซ) ไม่เหมาะสมกับข้อคำถามหลายข้อที่มีอำนาจจำแนกและค่าแลคซิมโกลต่ำหรือการ
เดาไม่เกินไปตามข้อตกลงเบื้องต้นของโมเดล นั่นคือโมเดลปฏิเสธข้อที่มีอำนาจจำแนก
สูงไปหรือต่ำไป และข้อที่มีโอกาสเดาถูกต้องจะนั้นจึงมีประโยชน์ในขั้นตอนของการพัฒนา
และเก็บรวบรวมข้อคำถามเฉพาะที่เหมาะสมกับโมเดลไว้ในธนาคารข้อสอบ

3. โมเดลดั้งเดิมให้ข้อมูลที่เป็ประโยชน์เฉพาะกับกลุ่มตัวอย่างที่ทดสอบ
เนื่องจาก ค่าประมาณพารามิเตอร์ของข้อขึ้นอยู่กับการกระจายของความสามารถของกลุ่ม
ผู้สอบด้วย นอกจากนี้การตรวจสอบรูปแบบการเลือกตอบข้อเลือกต่าง ๆ ยังเป็นประโยชน์
ในการเข้าใจกลุ่มนักเรียนอีกด้วยนอกเหนือจากการได้ข้อมูลเกี่ยวกับข้อคำถามแต่ละข้อ
โดยเฉพาะอย่างยิ่ง เมื่อเป็นแบบทดสอบที่ครูสร้างเอง ดังนั้นวิธีการวิเคราะห์รายข้อ

ตามแบบโมเดลดั้งเดิมจึงมีประโยชน์ในแง่เป็นข้อมูลประกอบกับผลจากการใช้โมเดล
โลจิสติก

ผู้ใช้โมเดลดั้งเดิมควรคำนึงไว้ว่าคะแนนที่ได้มาจากโมเดลนี้สัมพันธ์กับ
ชุดของคำถามที่ใช้และกลุ่มผู้สอบ คะแนนสูงอาจเนื่องมาจากใช้ชุดของข้อคำถามที่ง่าย ๆ
หรือสอบกับกลุ่มที่มีความสามารถต่ำ ๆ ก็ได้

4. การใช้โมเดลโลจิสติกในการวิเคราะห์รายชื่อต้องอาศัยคอมพิวเตอร์
ถึงแม้โมเดล 1PL จะมีวิธีคำนวณด้วยมือ (Wright, 1977) แต่ใช้เวลานานมาก
และสูตรที่ใช้มีความซับซ้อนพอสมควร สิ่งเหล่านี้อาจไม่สะดวกสำหรับโรงเรียนบางแห่ง
ในสถานการณ์ที่ไม่พร้อมในเรื่องดังกล่าวอาจวิเคราะห์โดยให้โมเดลดั้งเดิม ซึ่งให้
ค่าประมาณความยากที่มีอันดับเหมือนกันกับโมเดล 1PL ส่วนอำนาจจำแนกซึ่งขึ้นอยู่กับกลุ่ม
ผู้สอบนั้น สามารถแก้ปัญหาได้ด้วยวิธีการเลือกกลุ่มตัวอย่างอย่างระมัดระวังให้มีช่วงความ
สามารถที่คล้ายคลึงกับกลุ่มบุคคลที่จะใช้แบบทดสอบจริง ซึ่งเป็นข้อจำกัดของการใช้โมเดลนี้

เอกสารอ้างอิง

พรรคหทัย คงเสน่ห์ (2530) การสร้างแบบทดสอบสองขั้นตอน โดยให้ทฤษฎีการ

ตอบข้อคำถาม ปรินิพนธ์การศึกษามหาบัณฑิต บัณฑิตวิทยาลัย

มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร

จิราพร ไกรสรศิวเวท (2529) ความสัมพันธ์ระหว่างรูปแบบบางประการของแบบ

ทดสอบรูปปรามิตกับความสามารถทางการเรียนในวิชาคณิตศาสตร์ ปรินิพนธ์การ

ศึกษามหาบัณฑิต บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร

ัญญา แสนทวิ (2530) การประยุกต์รูปแบบของราส์ในการออกแบบโค้งแสดงสารสนเทศ

ของแบบสอบผลสัมฤทธิ์วิชาคณิตศาสตร์ตามระดับความสามารถของนักเรียนชั้นประถม

ศึกษาปีที่ 6 ปรินิพนธ์การศึกษามหาบัณฑิต ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย

จุฬาลงกรณ์มหาวิทยาลัย

บุญมี พันธุ์ไทย (2531) ผลของการรับรู้ผลกระทบของการสอบและแบบแผนการตอบที่มีต่อ

ความคลาดเคลื่อนมาตรฐานของการวัดของแบบสอบวัดผลสัมฤทธิ์วิชาคณิตศาสตร์

ชั้นมัธยมศึกษาปีที่ 4 ปรินิพนธ์การศึกษามหาบัณฑิต ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย

จุฬาลงกรณ์มหาวิทยาลัย

ผจงจิต อินทสุวรรณ (2528) ทฤษฎีการตอบข้อคำถาม สถาบันวิจัยพฤติกรรมศาสตร์

มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร

ธีระศักดิ์ คุรุจนาพันธ์ (2531) การเปรียบเทียบคุณภาพของแบบสอบแบบเลือกตอบ ที่มี

จำนวนตัวเลือกไม่เท่ากัน โดยให้ทฤษฎีการตอบสนองข้อสอบ ปรินิพนธ์การศึกษามหา

บัณฑิต ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

เรวดี อินทสระ (2530) การเปรียบเทียบความคลาดเคลื่อนของการเทียบมาตรา
ระหว่างรูปแบบอิงทฤษฎีการตอบข้อสอบกับรูปแบบการใช้เทคนิคการวิเคราะห์
องค์ประกอบ วิทยานิพนธ์ การศึกษามหาบัณฑิต บัณฑิตวิทยาลัย
มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร

สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี กระทรวงศึกษาธิการ (2532)
หนังสือเรียนคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 โรงพิมพ์คุรุสภาลาดพร้าว

สมจิตร ทรัพย์อัปราไมย (2531) การเปรียบเทียบผลการวิเคราะห์แบบทดสอบ
วัดเจตคติด้วยคลาสสิกัลโมเดลและราส์ชโมเดล วิทยานิพนธ์การศึกษามหาบัณฑิต
บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร

เสรี ประมวลลิขิตโรจน์ (2531) การศึกษาความเที่ยงตรงในการจำแนกความรอบรู้และ
การประมาณค่าความสามารถในวิชาคณิตศาสตร์ของแบบทดสอบแฟล็กชีเลเวล
และ แบบทดสอบรูปปิรามิดที่ใช้ขนาดชั้นแปรรัน วิทยานิพนธ์การศึกษามหาบัณฑิต
มหาวิทยาลัย ศรีนครินทรวิโรฒ ประสานมิตร

สุพัฒน์ สุกมลสันต์ (2530) การวิเคราะห์ข้อทดสอบแนวใหม่ด้วยเครื่องคอมพิวเตอร์
เอกสารลำดับที่ 1 ในโครงการธนาคารข้อสอบโดยใช้เครื่องคอมพิวเตอร์
สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย

สุรัตน์ ขวัญบุญจันทร์ (2531) การสร้างแบบทดสอบและตารางเทียบคะแนนในแนวนอน
วิชาวิทยาศาสตร์ที่วิเคราะห์ด้วยราส์ชโมเดล วิทยานิพนธ์การศึกษามหาบัณฑิต
บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร

ลาภรณ์ กาญจนกิจโสภณ (2531) การสร้างแบบทดสอบและตารางเก็บระดับคะแนน
ตามแนวคิดในวิชาคณิตศาสตร์ ปรินญาการศึกษามหาบัณฑิต บัณฑิตวิทยาลัย
มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร

Anastasi, A. (1976). Psychological testing (Fourth edition).
New York: Macmillan.

Andersen, E.B. (1970). Asymptotic properties of conditional
maximum likelihood estimates. Journal of the Royal
Statistical Society, 32, 283-301.

Baker, F.B. (1977). Advances in item analysis. Review of
Educational Research, 47(1), 151-178.

Bejar, I.I. (1980). A procedure for investigating the
unidimensionality of achievement tests based on item
parameter estimates. Journal of Educational Measurement,
17(4), 283-296.

Birnbaum, A. (1968). Some latent trait models. In Lord, F.M.,
and Novick, M.R. Statistical theories of mental test
scores. Massachusetts: Addison-Wesley.

Bloom, H.S., Hastings, J.T., & Madaus, G.F. (1971). Handbook
on formative and summative evaluation of student learning.
New York: McGraw-Hill.

Cronbach, L.J. (1970). Essentials of psychological testing

(Third edition). New York: Harper & Row.

Davis, F.B. (1974). (chairman of the joint committee) Standards

for educational and psychological test. Washington, D.C.:

American Psychological Association.

Dixon, W.J., & Brown, M.B. (Eds.), (1979). BMDP-79. Los Angeles:

University of California Press.

Dragow, F., & Parsons, C.K. (1983). Application of

unidimensional item response theory models to multidimensional

data. Applied Psychological Measurement, 7(2), 189-199.

Green, S.B., Lissitz, R.W., & Mulaik, S.A. (1977). Limitations

of coefficient alpha as an index of test unidimensionality.

Educational and Psychological Measurement, 37, 827-838.

Guilford, J.P. (1954). Psychometric methods. New York: McGraw-

Hill Book.

Gulliksen, H. (1950). Theory of mental tests. New York: Wiley.

Hambleton, R.K., & Swaminathan, H. (1985). Item response theory:

Principles and applications. Boston: Kluwer-Nijhoff.

Hambleton, R.K., & Rovinelli, R.J. (1986). Assessing the

dimensionality of a set of test items. Applied Psychological

Measurement, 10(3), 287-302.

- Henrysson, S.(1976). Gathering, analysing, and using data on test items. In Thorndike, R.L.(Ed.), Educational measurement. Washington, D.C.:American Council on Education.
- Hinkle, D.E., Wiersma, W., & Jurs, S.G.(1979). Applied statistics for the behavioral sciences. Boston: Houghton Mifflin.
- Jensen, C.(1976) A simple technique for estimating latent trait mental test parameters. Educational and Psychological Measurement, 36, 705-715.
- Keats, J.A.(1989). Personal Communication.
- Kingston, N.(1986). Assessing the dimensionality of the GMAT verbal and quantitative measures using full information factor analysis. Research Report 86-13. Princeton, NJ: Educational Testing Service.
- Kline, P.(1986). A handbook of test construction: Introduction to psychometric design. London: Methuen.
- Lord, F.M.(1968). An analysis of the verbal scholastic aptitude test using Birnbaum's three parameter logistic model. Educational and Psychological Measurement, 28, 989-1020.
- Lord, F.M.(1974). Estimation of latent ability and item parameters when there are omitted responses. Psychometrika, 39(2), 247-264.

- Lord, F.M. (1980). Applications of item response theory to practical testing problems. New Jersey: Lawrence Erlbaum Associates.
- Lord, F.M., & Novick, M.R. (1968). Statistical theories of mental test scores. Massachusetts: Addison Wesley.
- Masters, G. (1984). DICOT: Analyzing classroom tests with the Rasch model. Educational and Psychological Measurement, 44, 145-150.
- McDonald, R.P. (1981). The dimensionality of tests and items. British Journal of Mathematical and Statistical Psychology 34, 100-117.
- Nie, N.H. (1983). SPSS^x user's guide. Illinois: SPSS.
- Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen: Denmark's Paedagogiske Institute.
- Reckase, M.D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. Journal of Educational Statistics, 4(3), 207-230.
- Samejima, F. (1973). A comment of Birnbaum's three parameter logistic model in the latent trait theory. Psychometrika, 38, 221-233.

SAS user's guide: Statistics version 5 edition (1985). North Carolina: SAS Institute.

Smith, J.K. (1980). On the examination of test unidimensionality.

Educational and Psychological Measurement, 40, 885-889.

Wainer, H. (1989). The future of item analysis. Journal of

Educational Measurement, 26(2), 191-208.

Wingersky, M.S. Barton, M.A., & Lord, F.M. (1982). LOGIST user's guide. Princeton, N.J.: Educational Testing Service.

Wood, R.L., Wingersky, M.S., & Lord, F.M. (1976). LOGIST: A computer program for estimating examinee ability and item characteristic curve parameters. Research Memorandum (modified 1/78), Educational Testing Service, Princeton, New Jersey.

Wright, B.D. (1977). Solving measurement problems with the Rasch model. Journal of Educational Measurement, 14(2), 97-116.

Wright, B.D., & Bell, S.R. (1977). Verifying the unconditional estimation procedure for Rasch item analysis with simulated data. (RM 26) Statistical Laboratory, Department of Education, University of Chicago.

- Wright, B.D., & Douglas, G.A. (1977). Conditional versus unconditional procedures for sample-free item analysis. Educational and Psychological Measurement, 37, 573-586.
- Wright, B.D., Mead, R.J., & Bell, S.R. (1980). BICAL: Calibrating items with the Rasch model. Research Memorandum 23 C, Chicago. Statistical Laboratory, Department of Education, University of Chicago.
- Wright, B.D. & Panchapakesan N. (1969). A procedure for sample-free item analysis. Educational and Psychological Measurement, 29, 23-48.
- Wright, B.D., & Stone, M.H. (1979). Best test design: Rasch measurement. Chicago: MESA Press.
- Yen, W.M. (1984). Obtaining maximum likelihood trait estimates from number-correct scores for the three-parameter logistic model. Journal of Educational Measurement, 21(2), 93-111.

ภาคผนวก ก

แสดงหมายเลขข้อในแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิทยาศาสตร์

ชั้นประถมศึกษาปีที่ 6 จำแนกตามเนื้อหาและพฤติกรรมต่าง ๆ

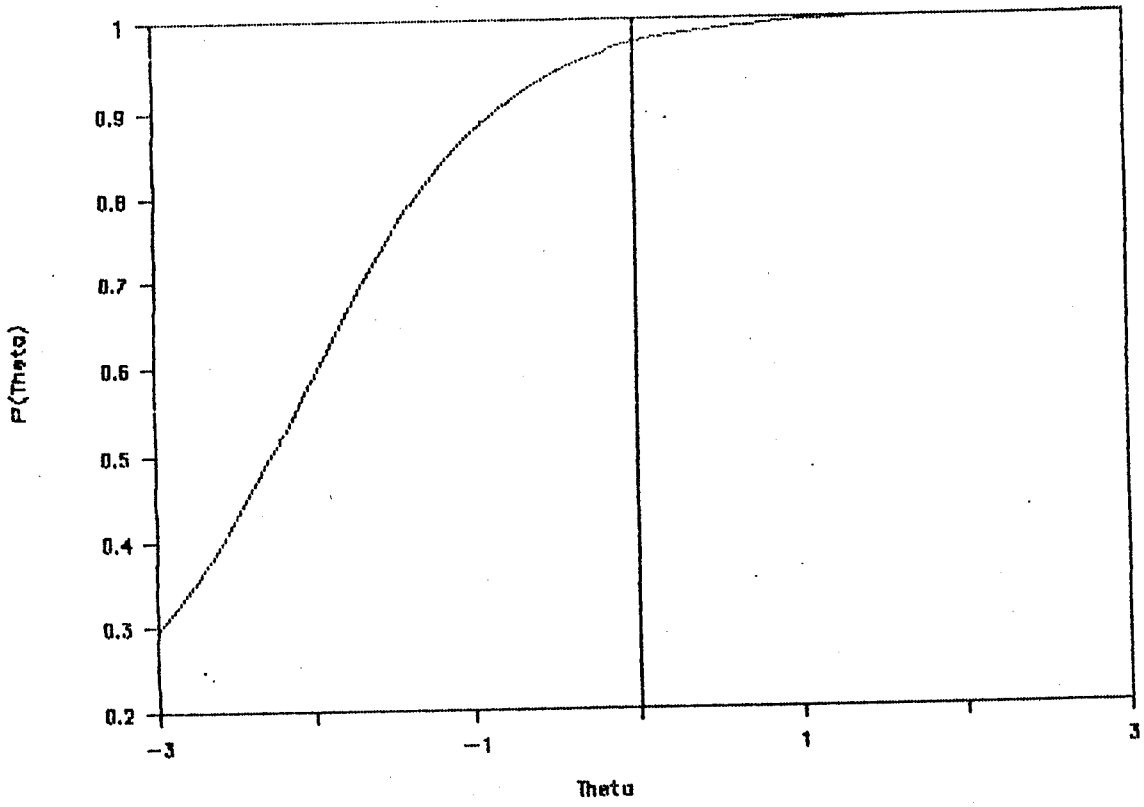
เนื้อหา	หมายเลขข้อที่วัดพฤติกรรม				รวม
	ความจำ	ความเข้าใจ	การนำไปใช้	การวิเคราะห์	
1. จำนวนและตัวเลข	1-5	-	-	-	5
2. สมการ	20	21-23	54-56	-	7
3. ตัวประกอบของจำนวนนับ	24, 25	-	-	-	2
4. ความเท่ากันทุกประการ	-	-	-	12	1
5. รูปสมมาตร	-	-	-	13, 14	2
6. มุมและส่วนของเส้นตรง	6, 7	-	-	-	2
7. เศษส่วน	-	-	42-44	-	3
8. ทศนิยม	-	-	47	-	1
9. ทศนิยม	-	-	45	-	1
10. เส้นขนาน	-	-	8, 9	-	2
11. รูปสามเหลี่ยม	-	-	-	-	0
12. บทประยุกต์	-	-	30, 31-35, 46	-	7
13. ทิศและแผนผัง	26, 29, 59	28, 40, 57, 58, 60	27	-	9
14. รูปสี่เหลี่ยม	-	-	10	-	1

เนื้อหา	หมายเลขที่วัดพฤติกรรม				รวม
	ความจำ	ความเข้าใจ	การนำไปใช้	การวิเคราะห์	
15. ความยาวรอบรูปและ					
พื้นที่ของรูปเหลี่ยม	11	41	36, 38, 50, 51	-	6
16. ความยาวรอบรูปและพื้นที่					
ของรูปวงกลม	-	-	37, 39	-	2
17. รูปทรงและปริมาตร	-	-	48, 49	15, 16	4
18. คู่อันดับ	-	-	19	-	1
19. แผนภูมิและกราฟ	-	17, 18, 52	53	-	4

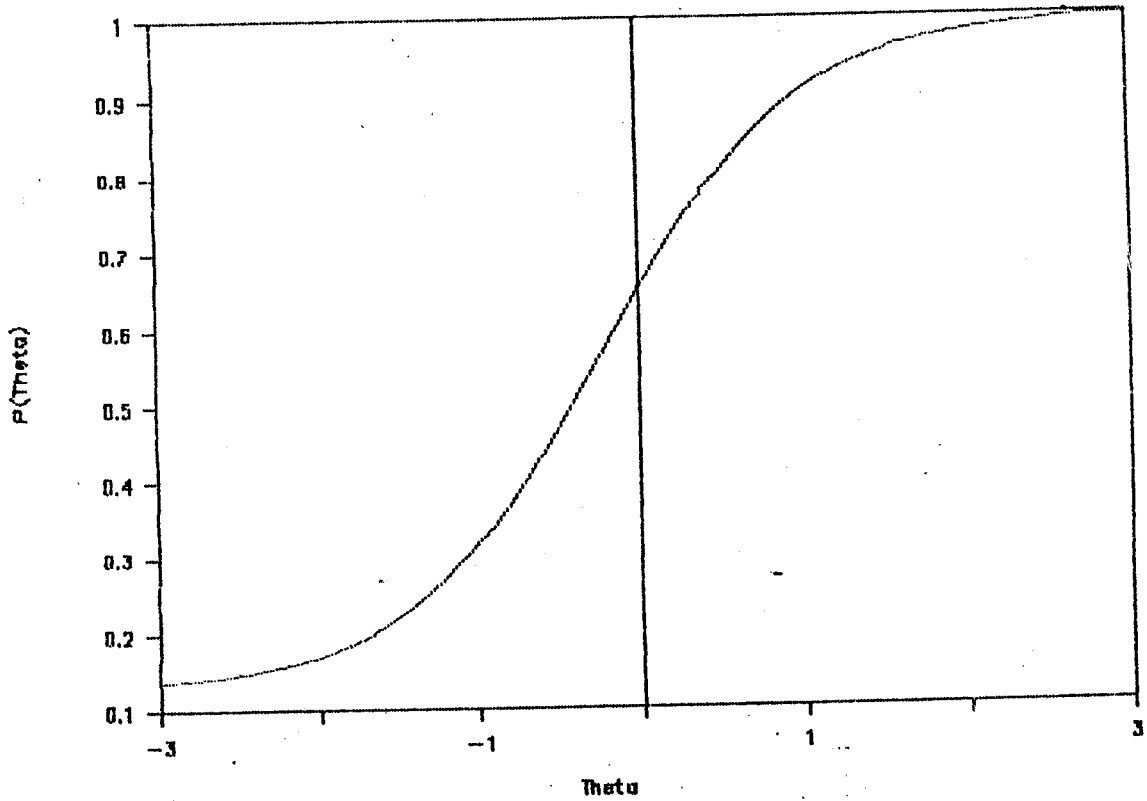
ภาคผนวก ข

โค้งแสดงลักษณะของข้อคำถาม 59 ที่อ ในแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน
วิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ซึ่งประมาณค่าพารามิเตอร์ภายใต้โมเดล 3PL

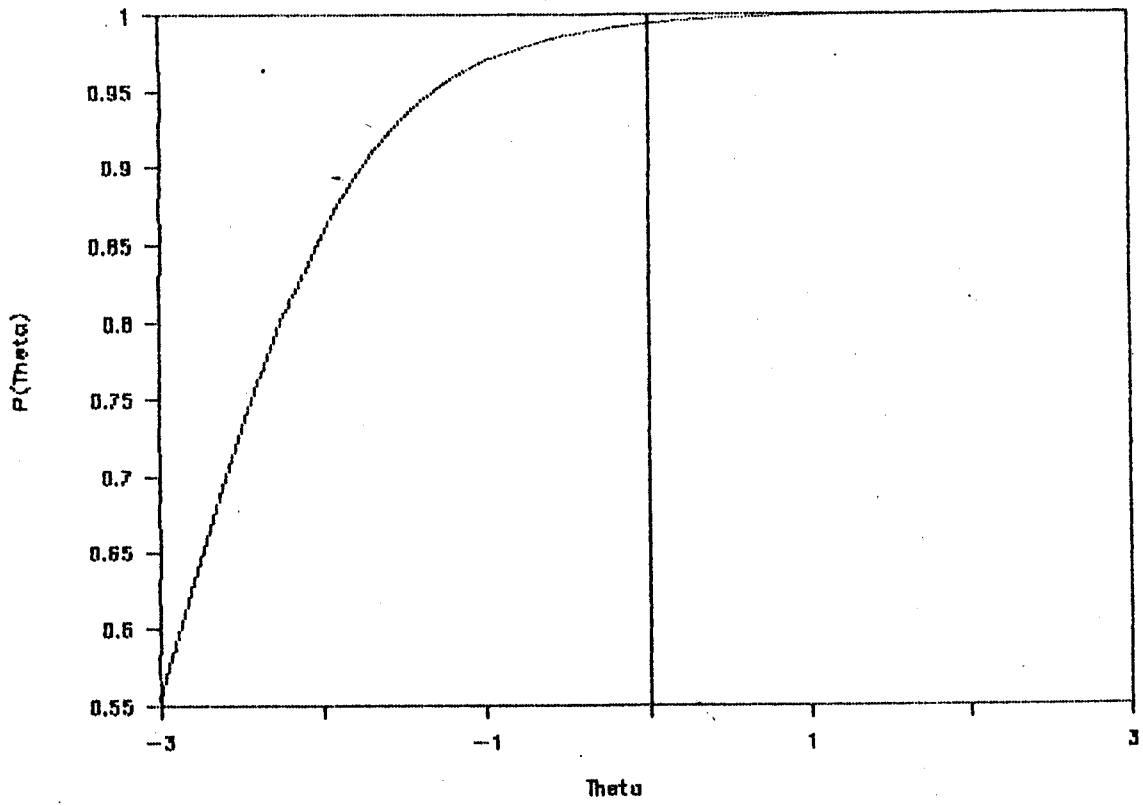
ITEM Q1



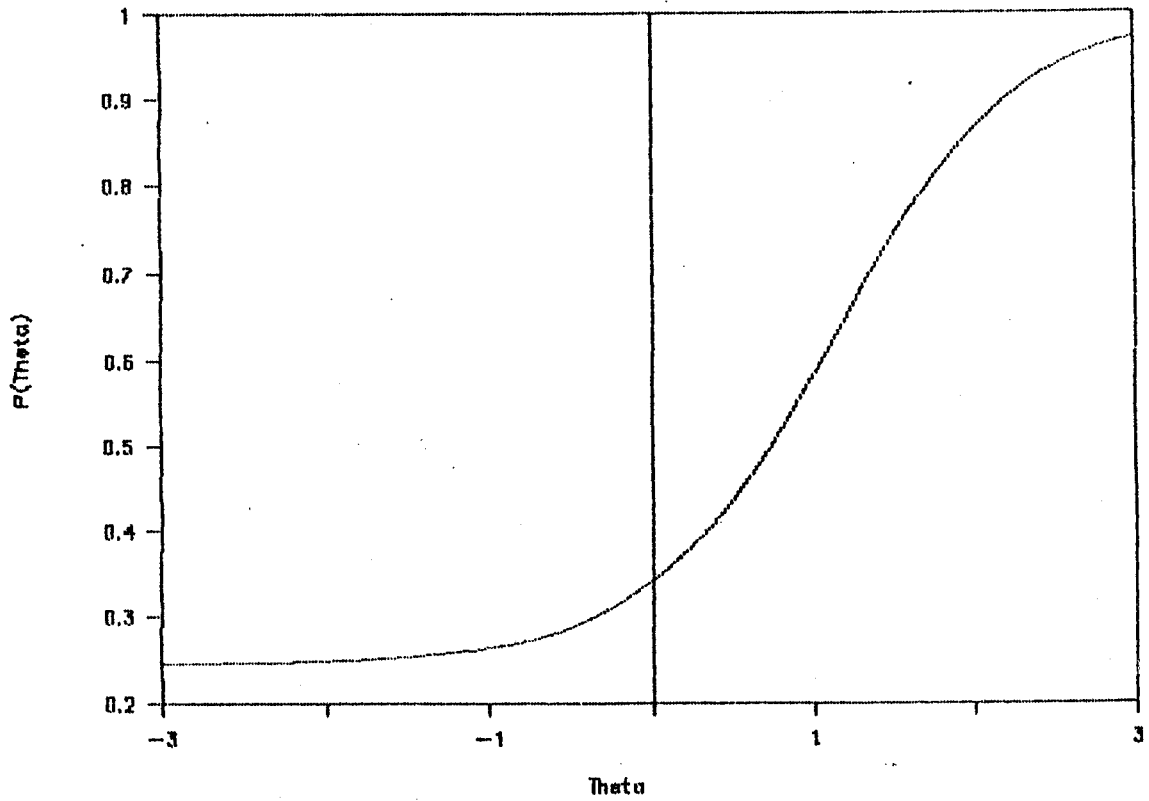
ITEM Q2



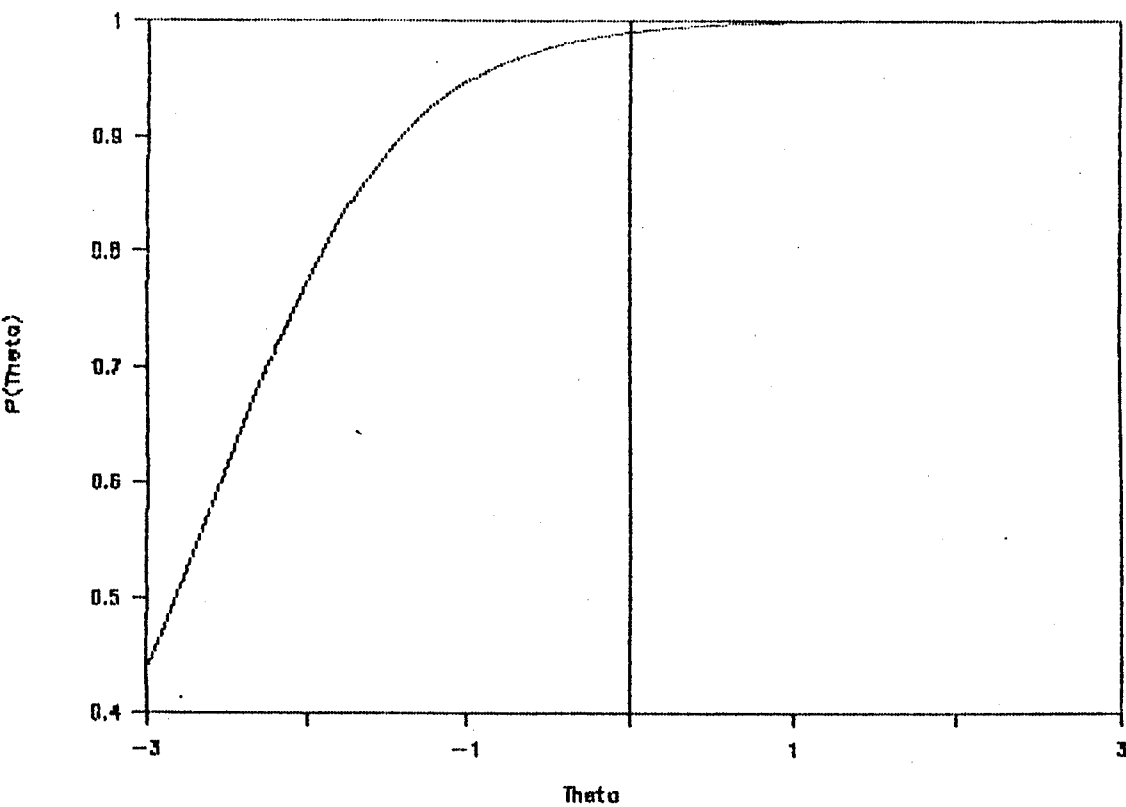
ITEM 03



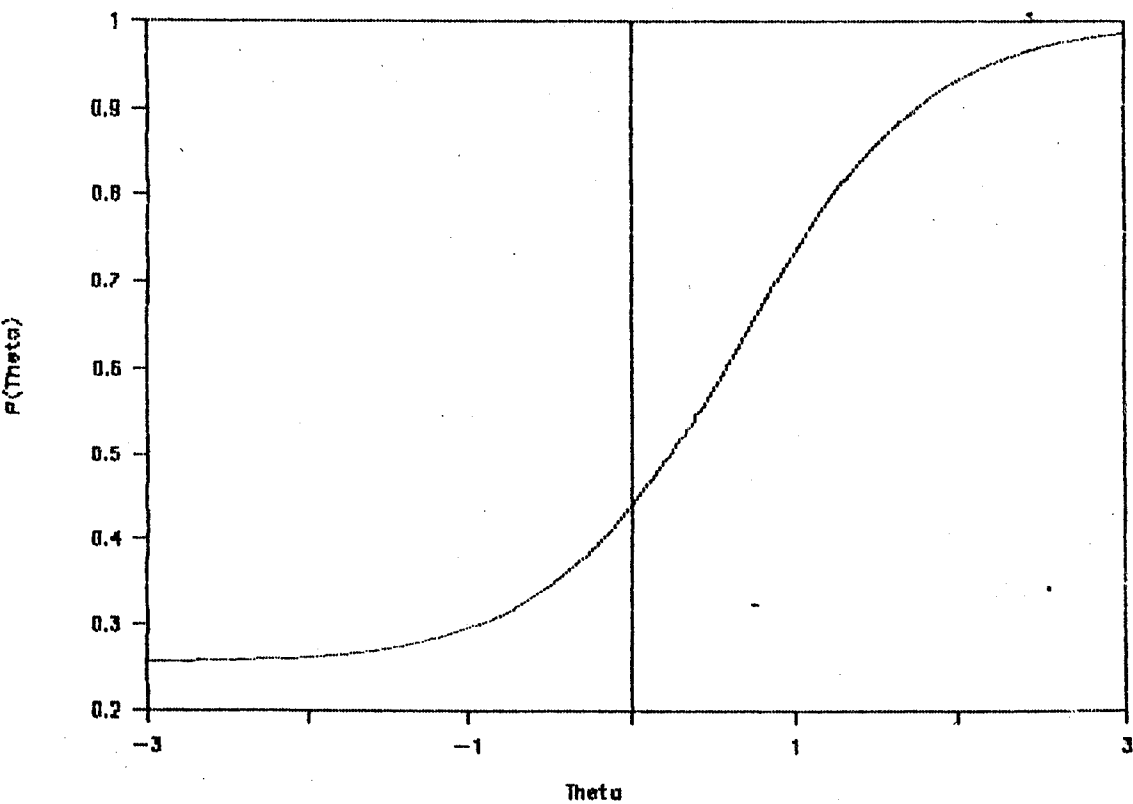
ITEM 04



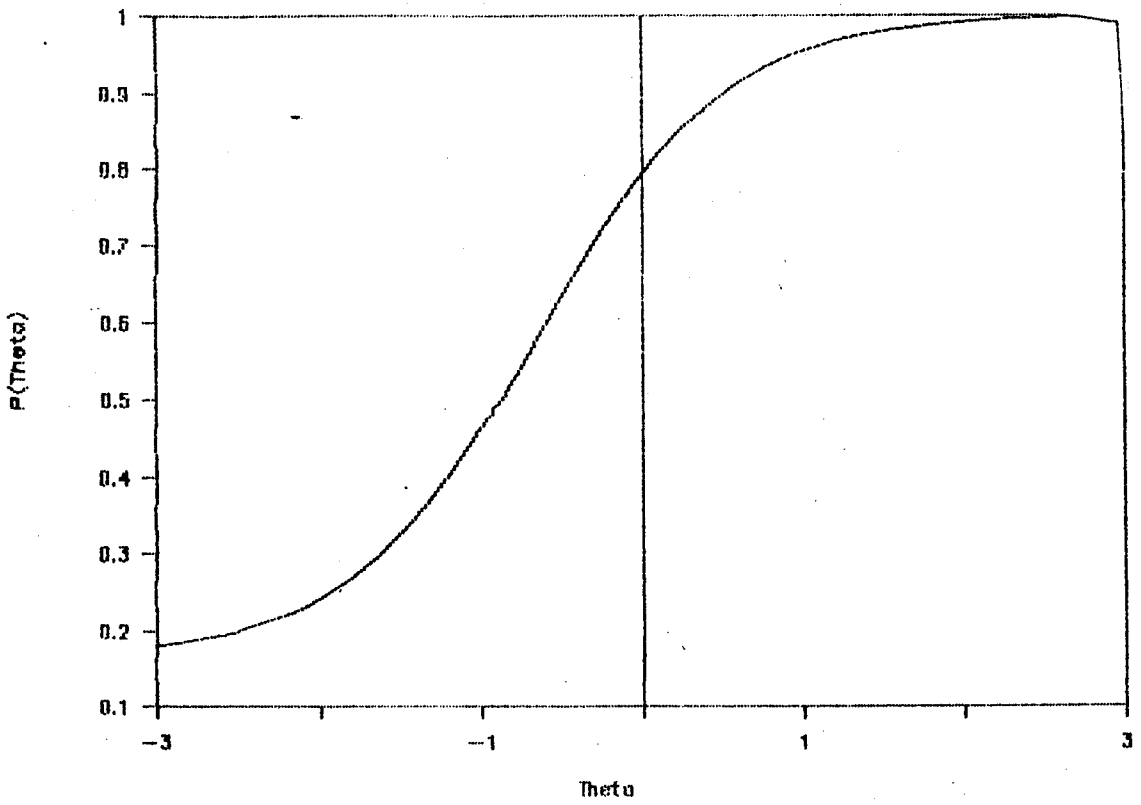
ITEM 09



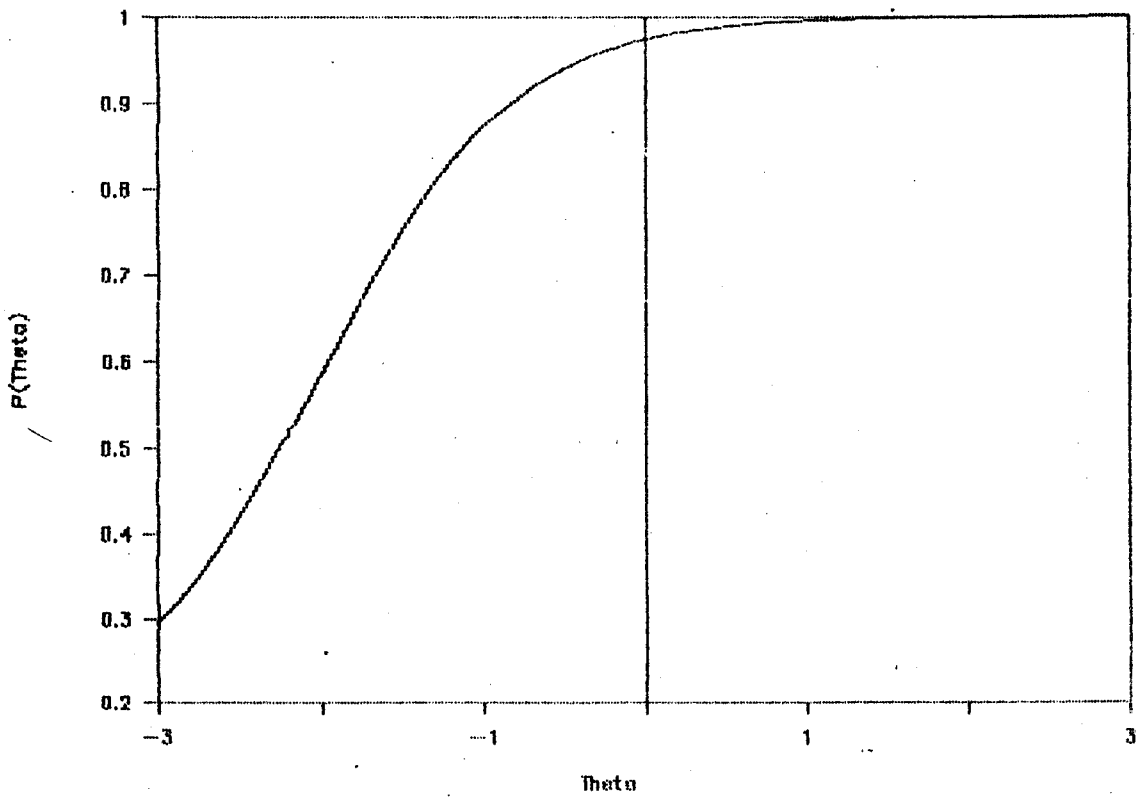
ITEM 10



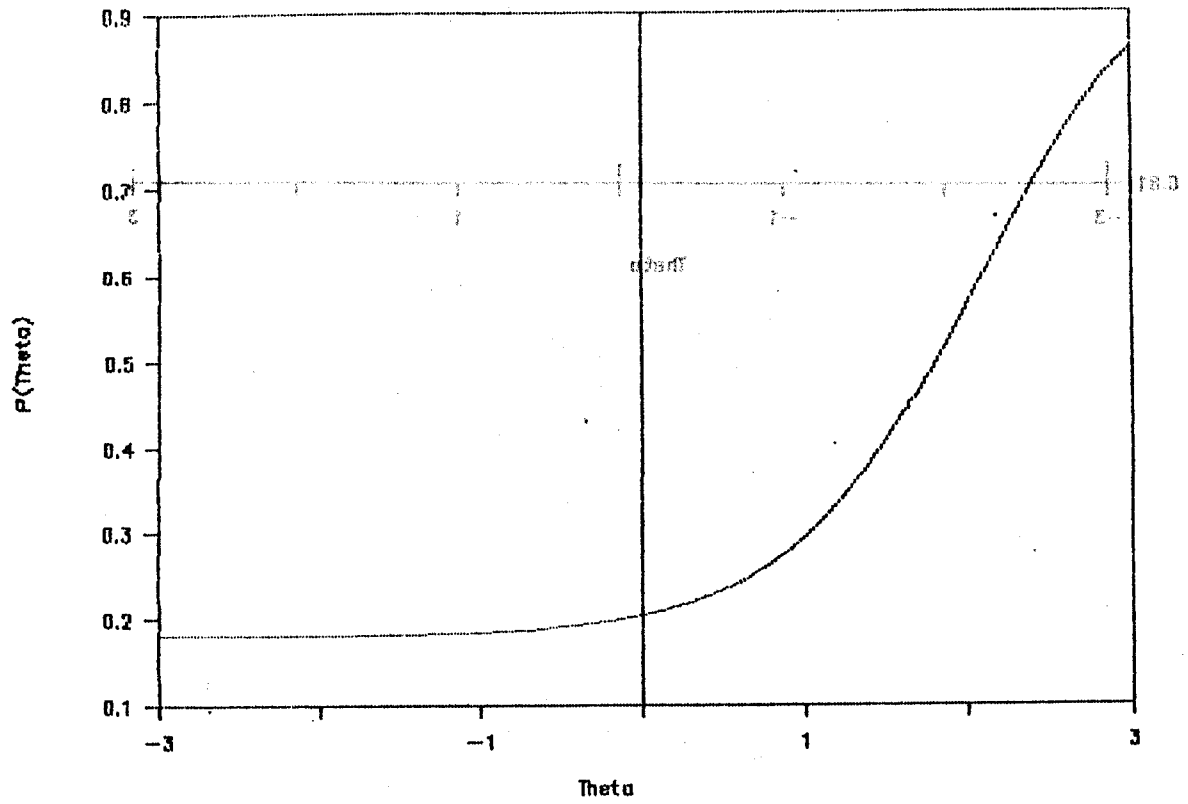
ITEM11



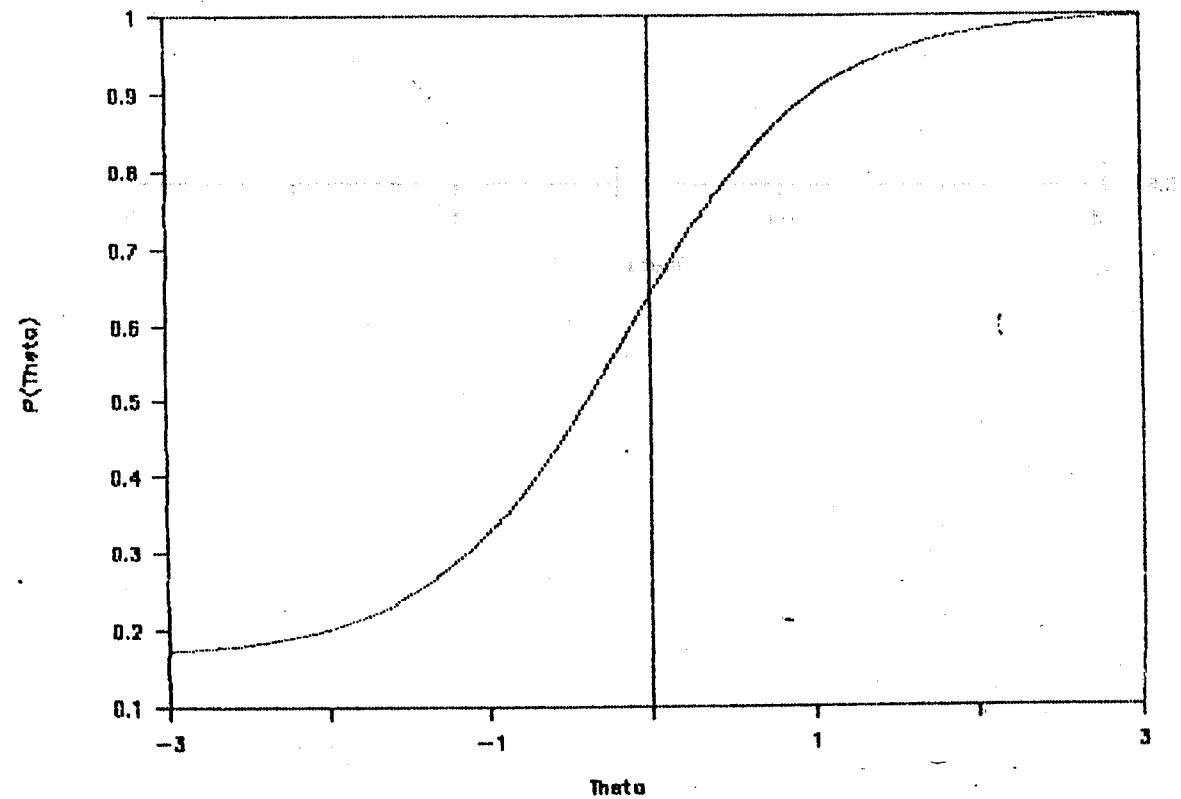
ITEM12



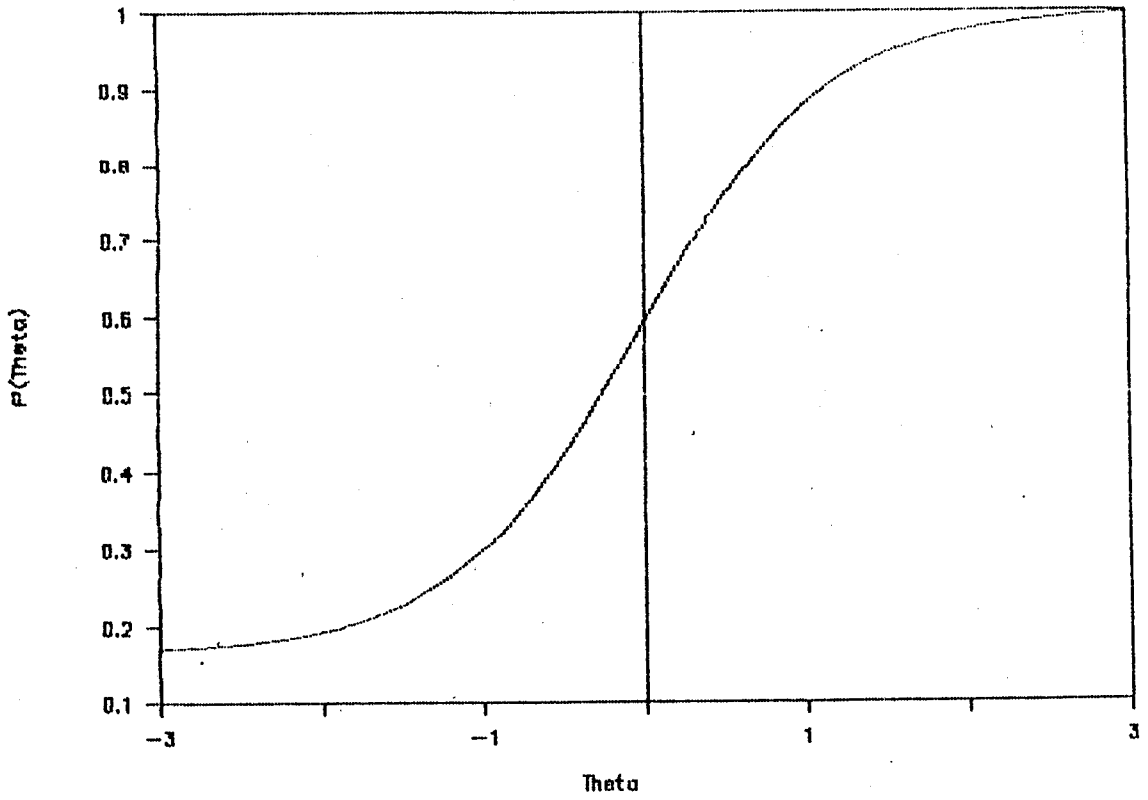
ITEM13



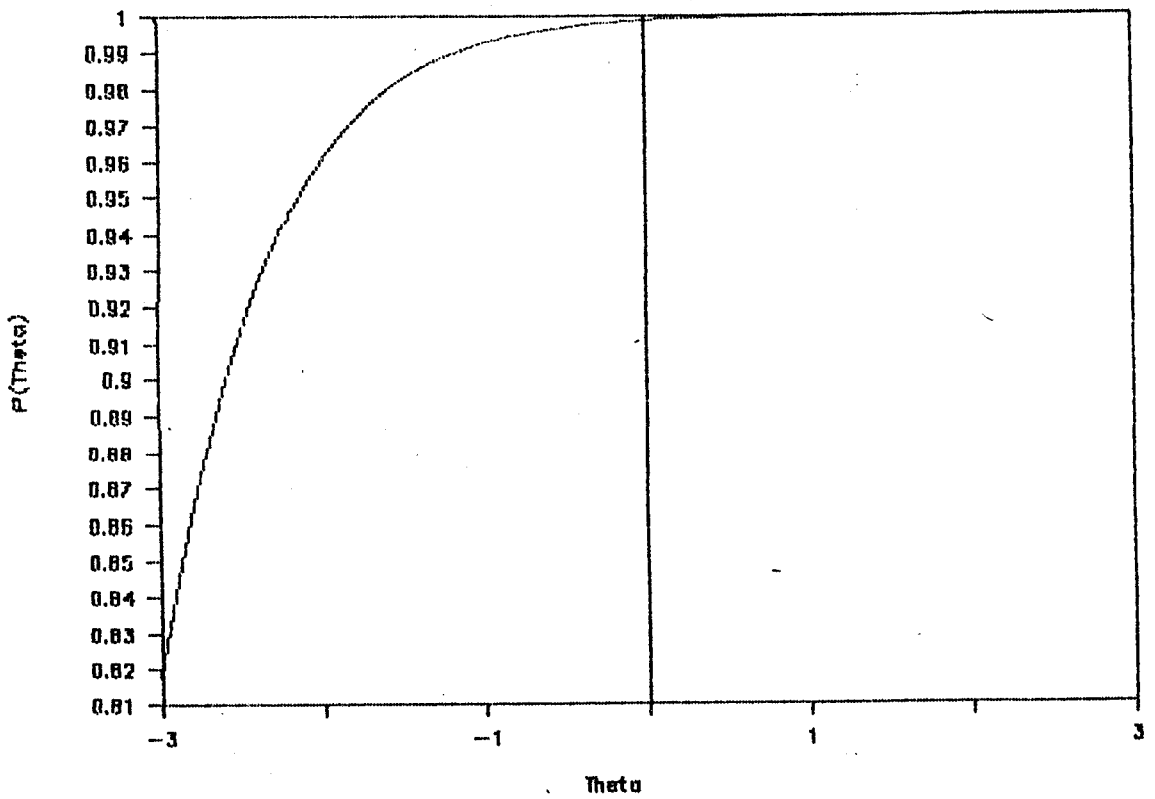
ITEM15



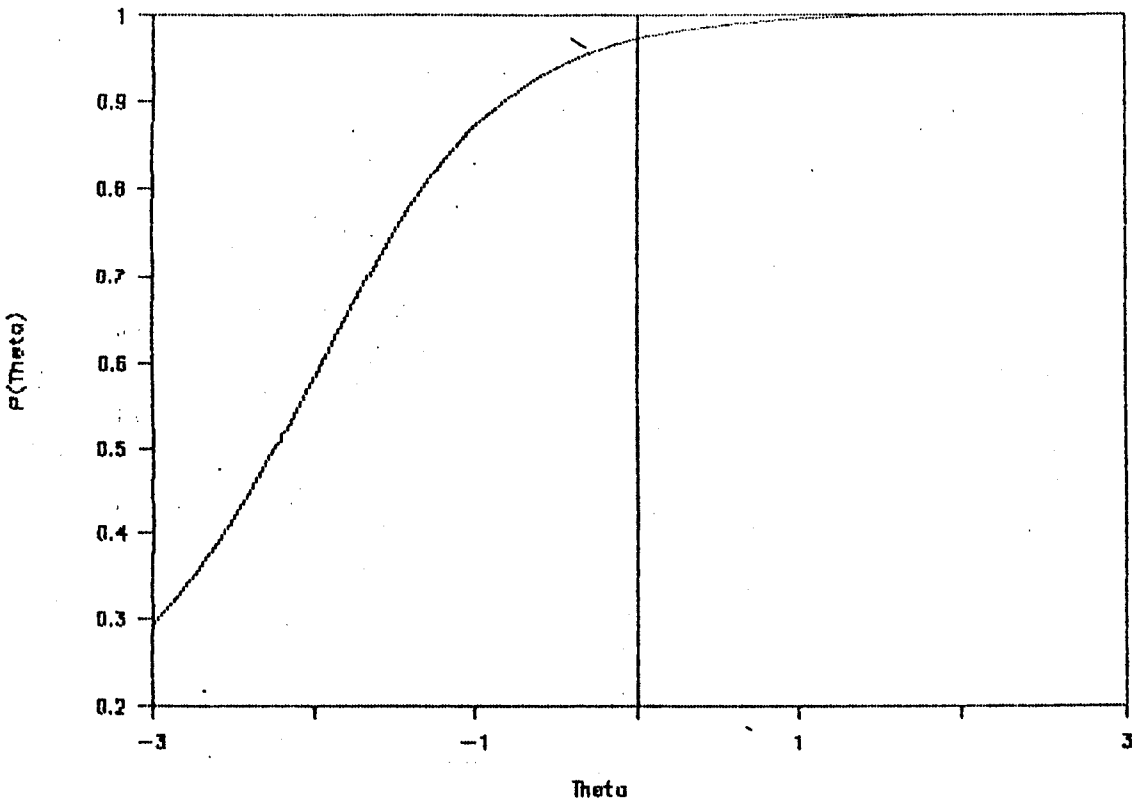
ITEM16



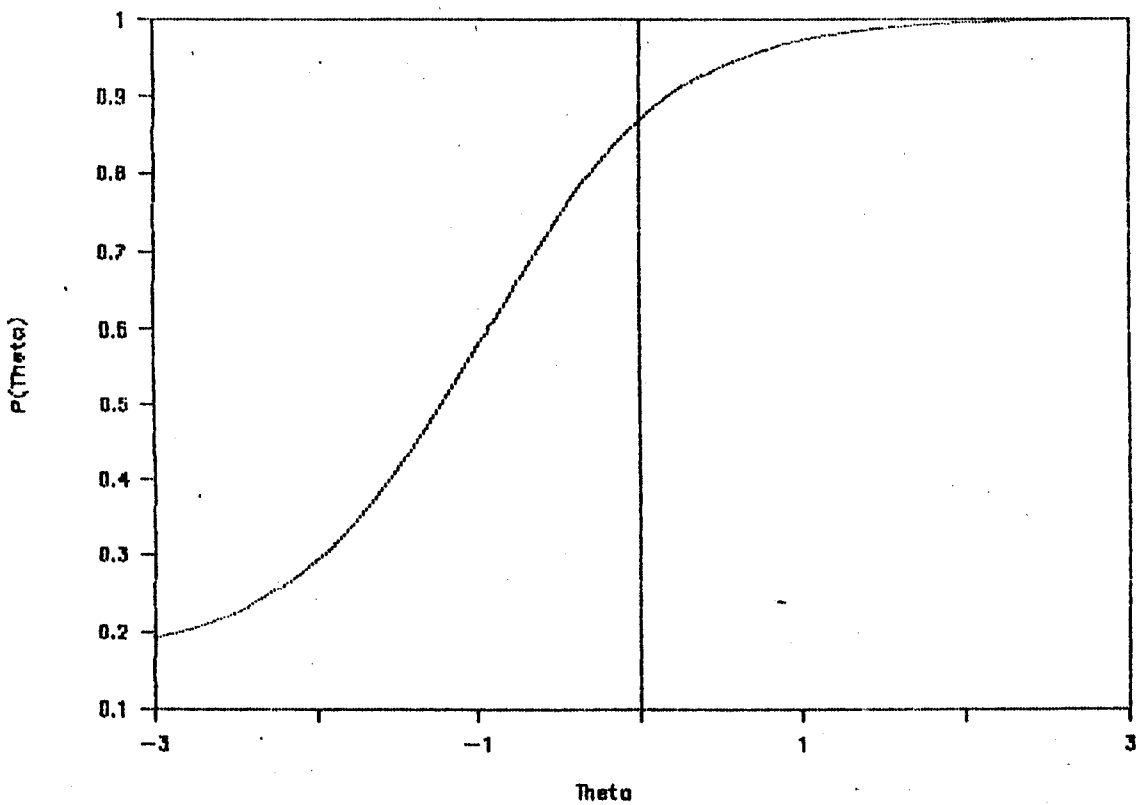
ITEM17



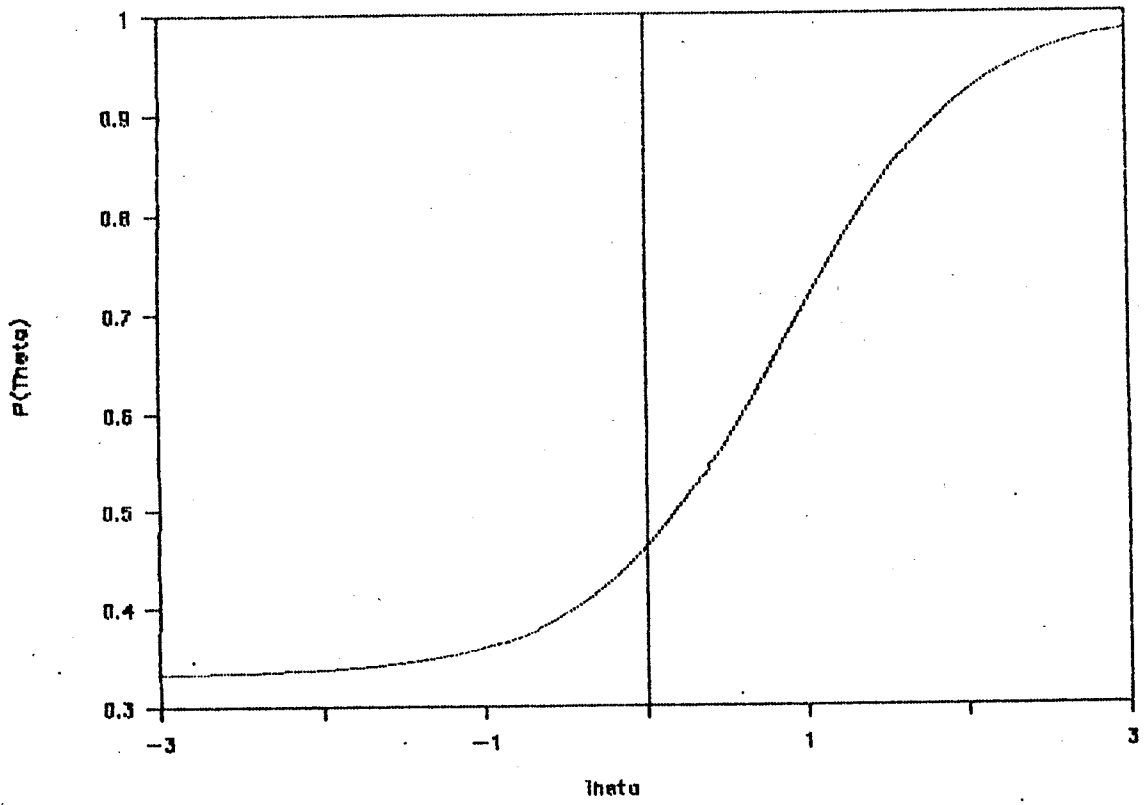
ITEM18



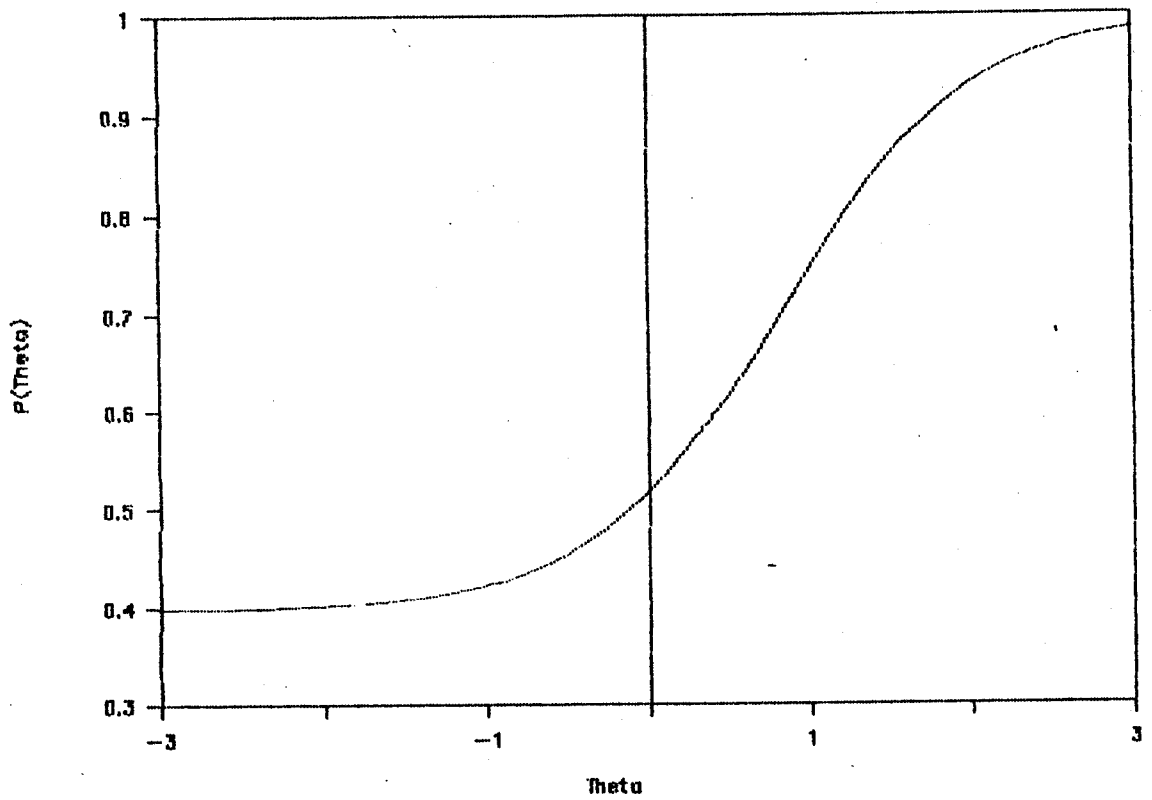
ITEM19



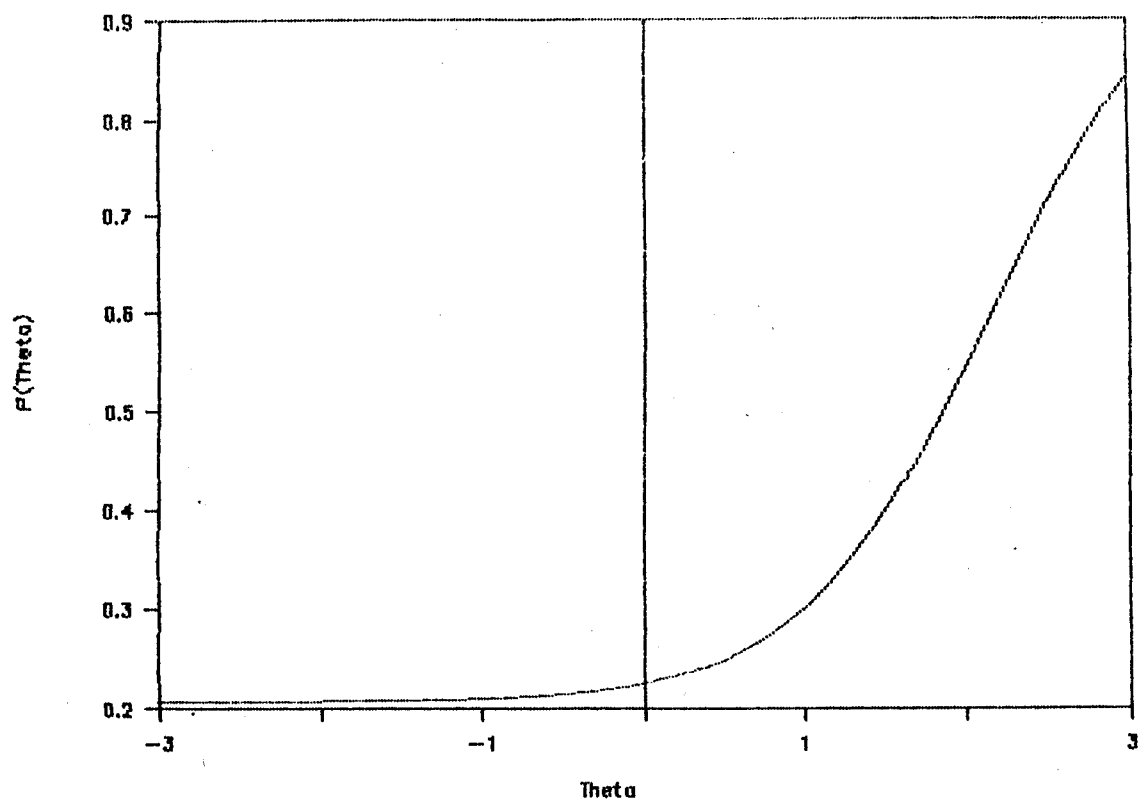
ITEM20



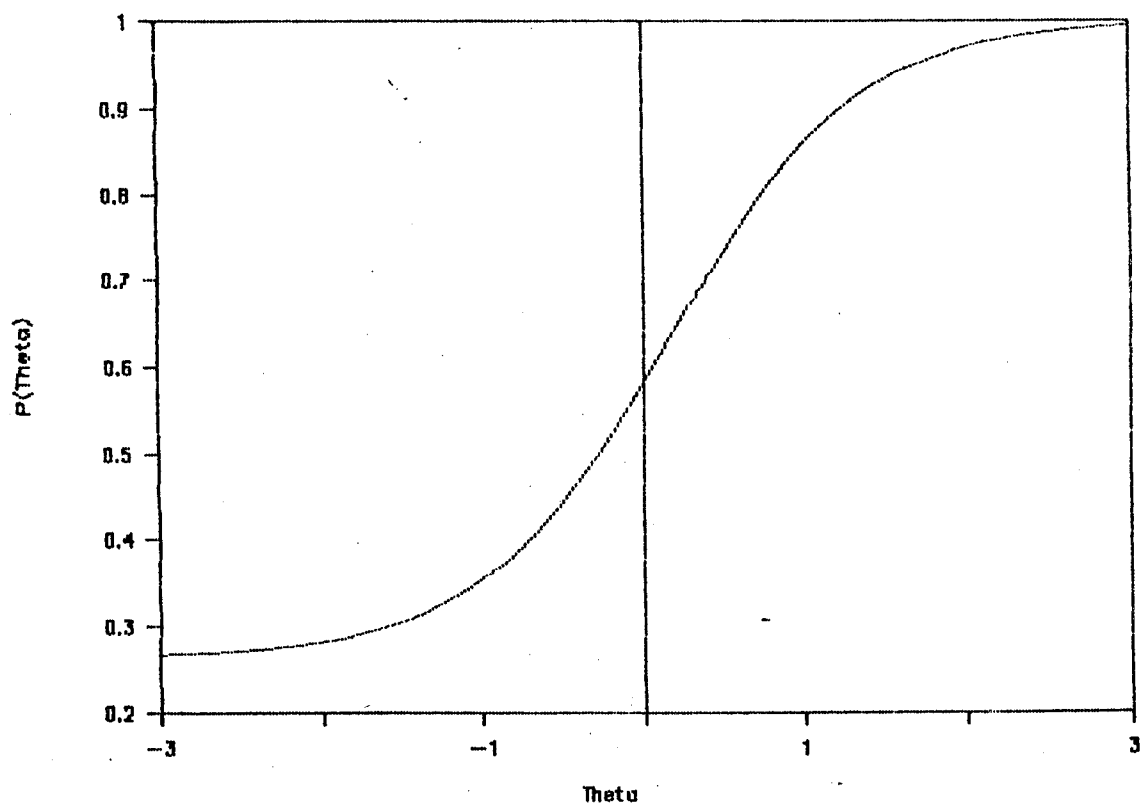
ITEM21



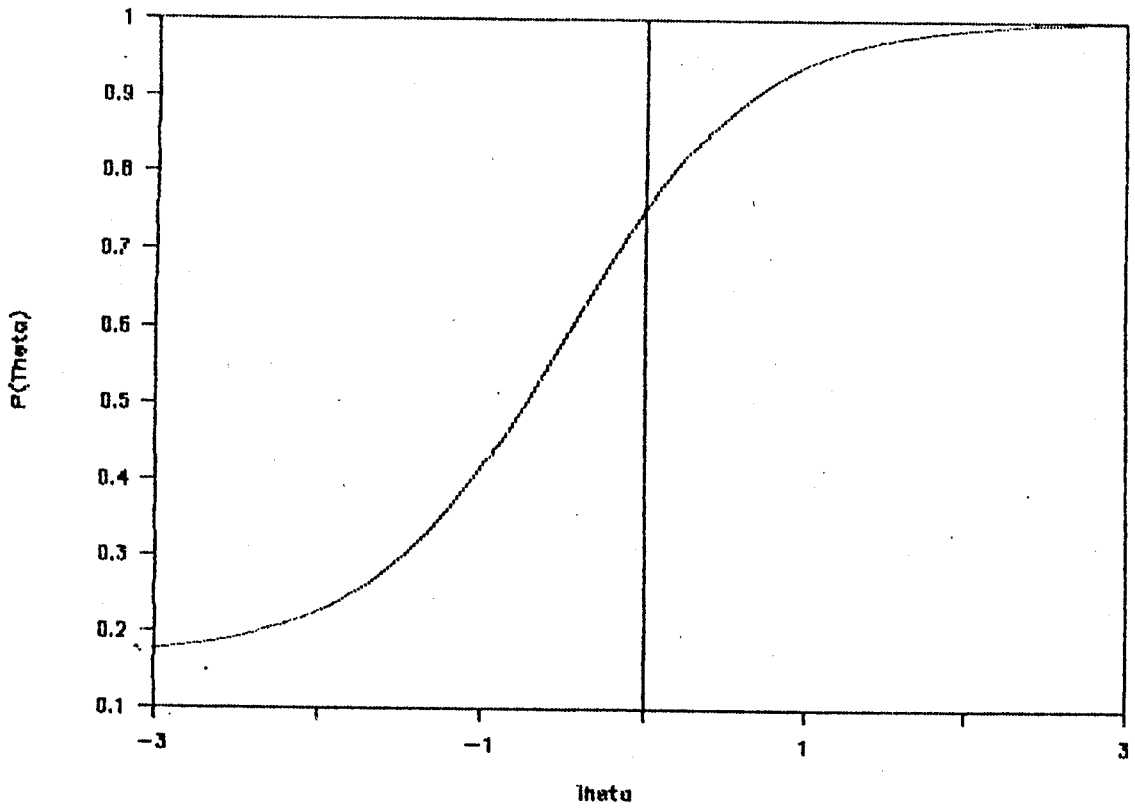
ITEM 22



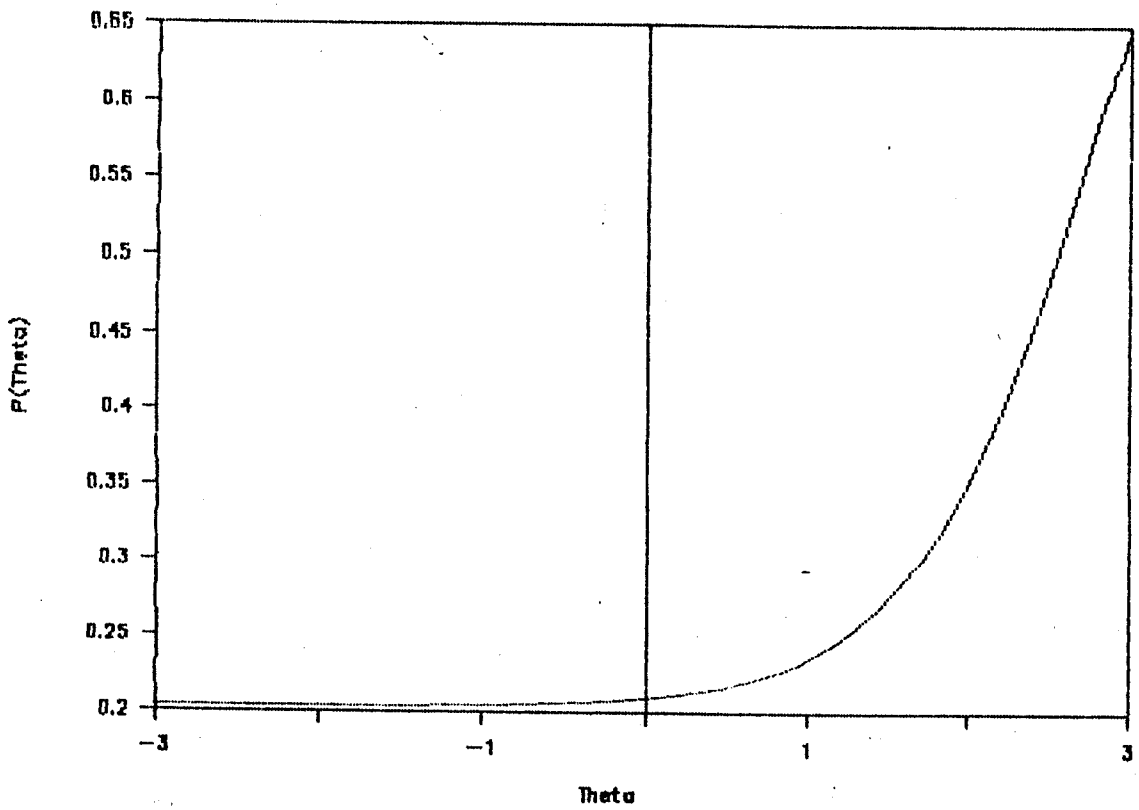
ITEM 23



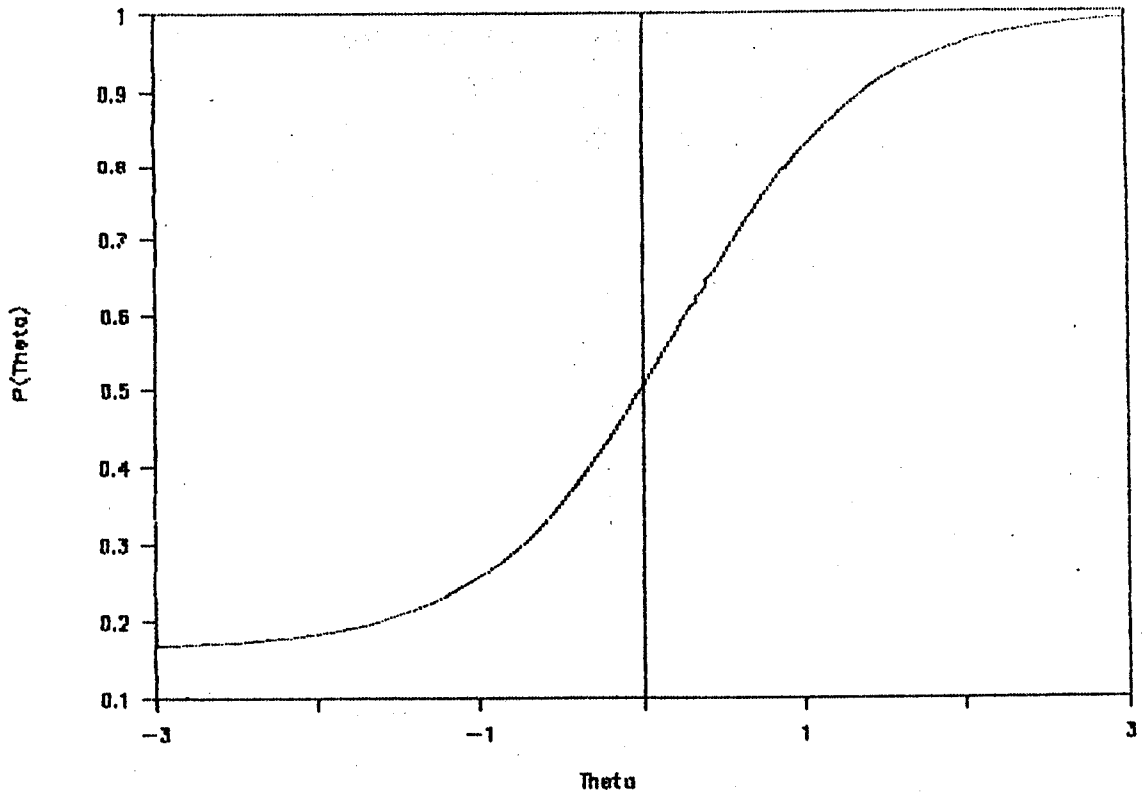
ITEM 24



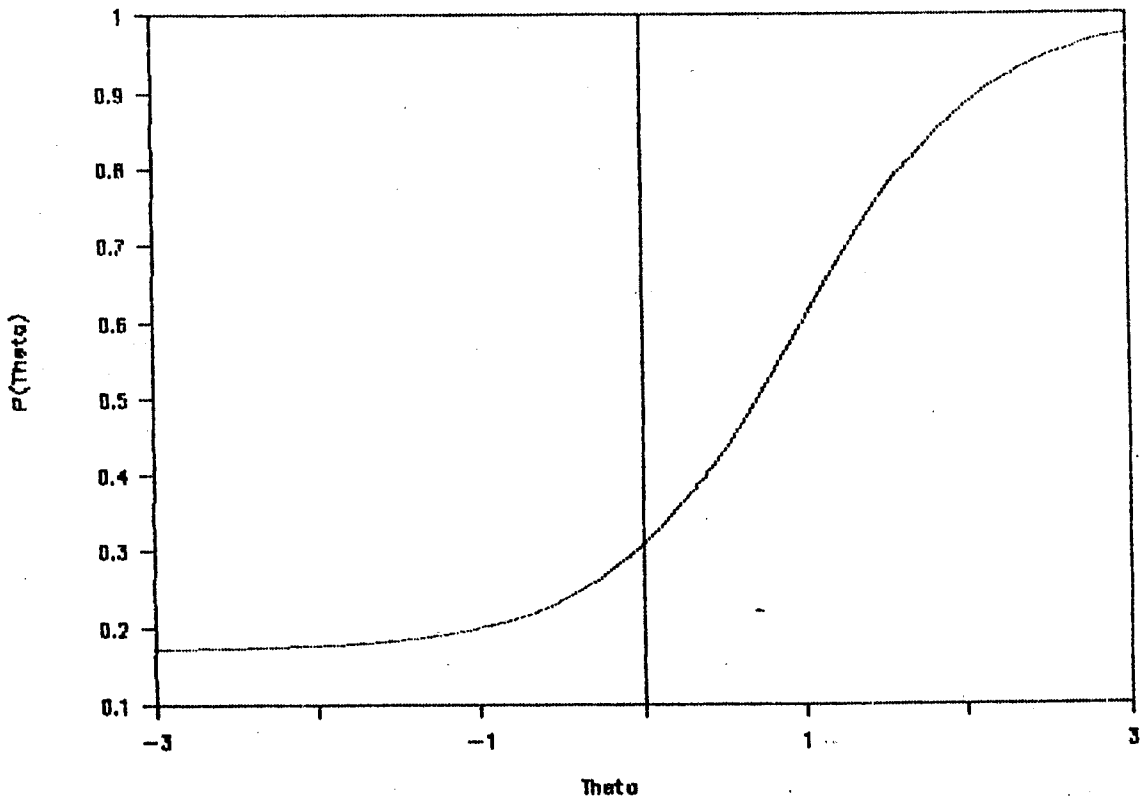
ITEM 25



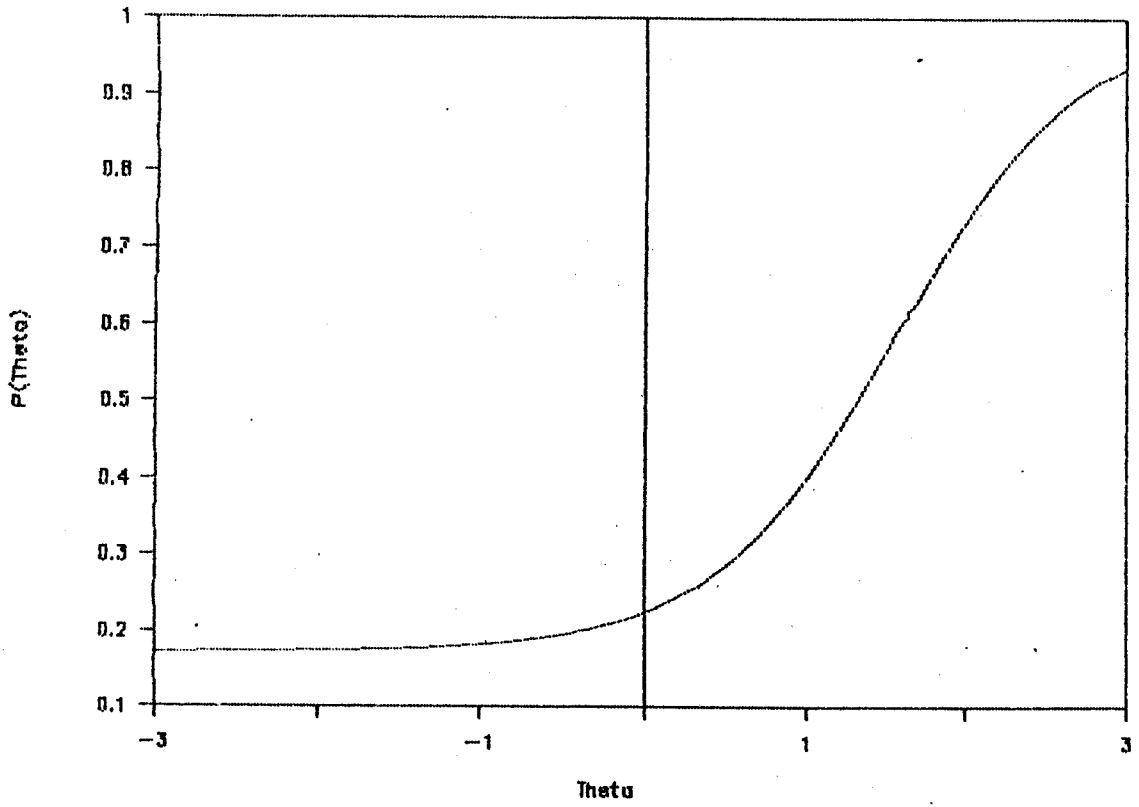
ITEM 26



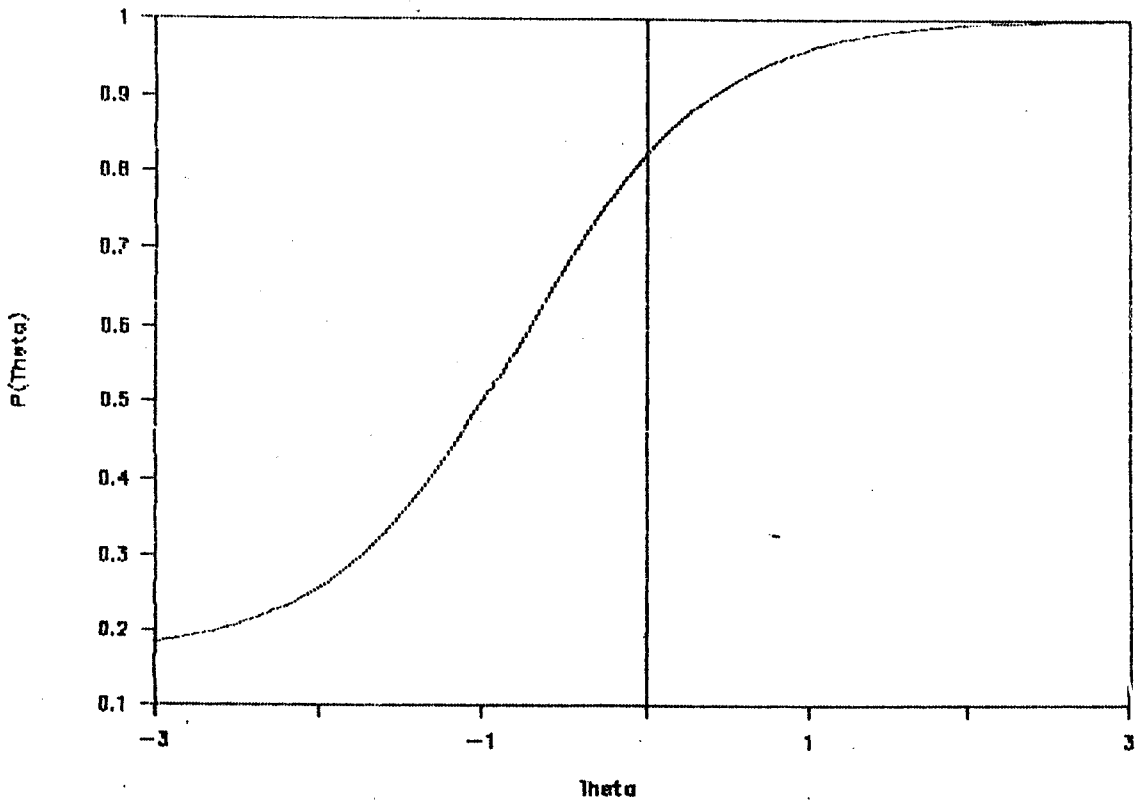
ITEM 27



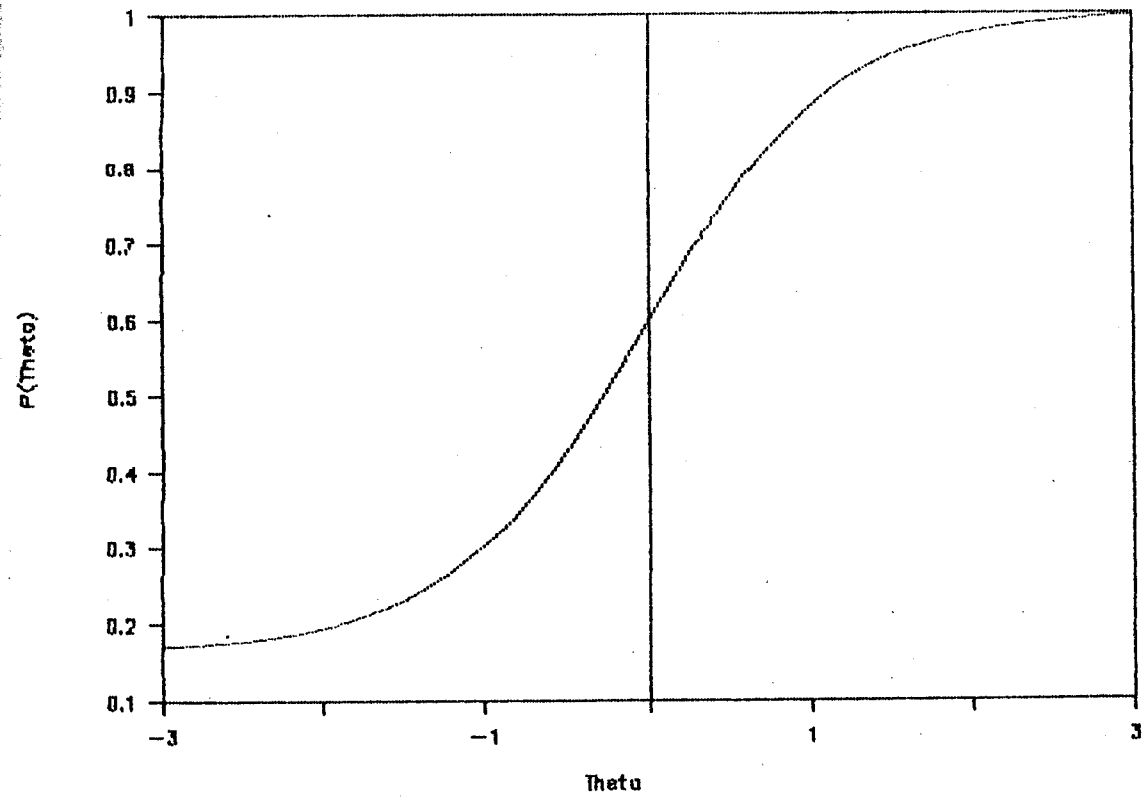
ITEM 28



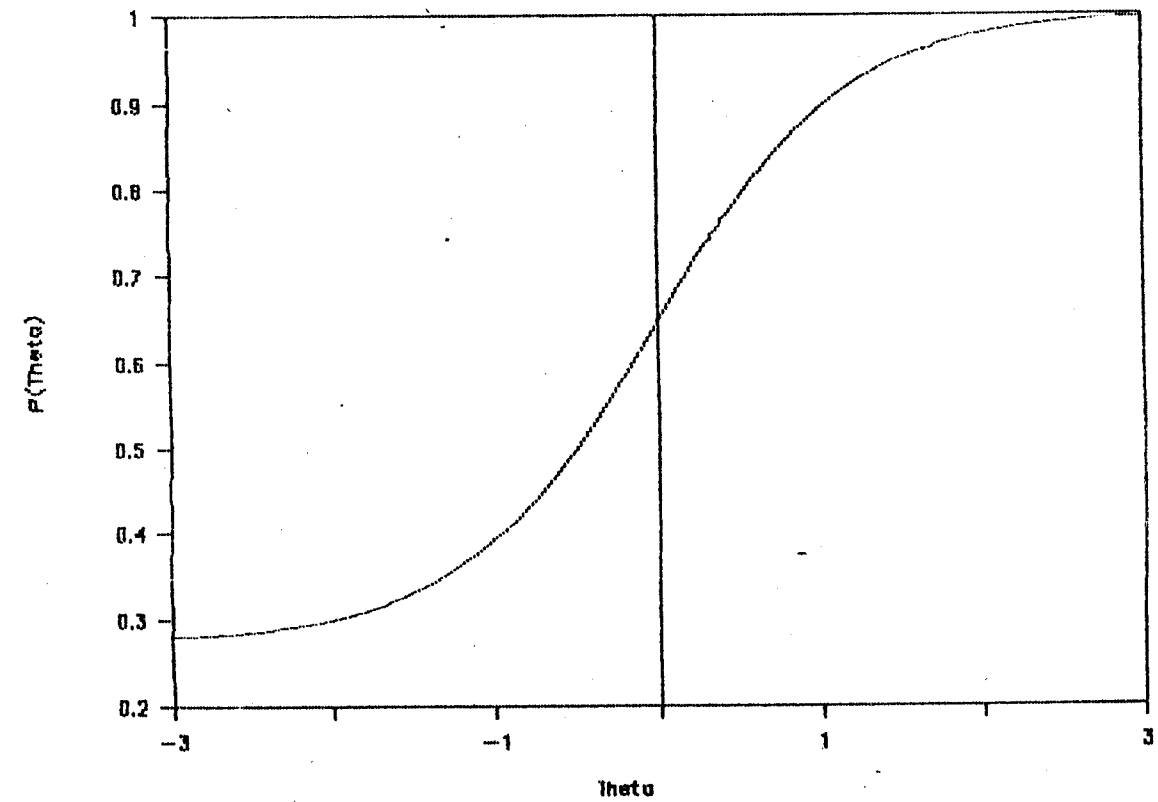
ITEM 29



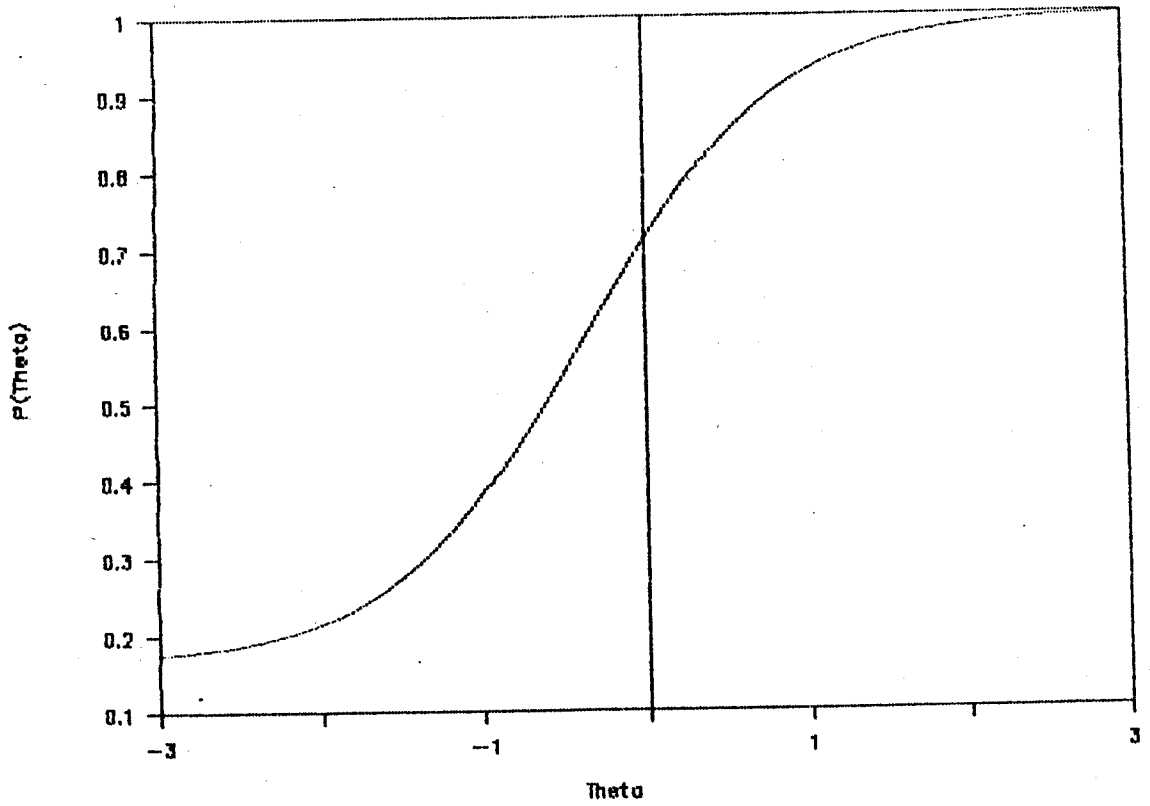
ITEM 30.



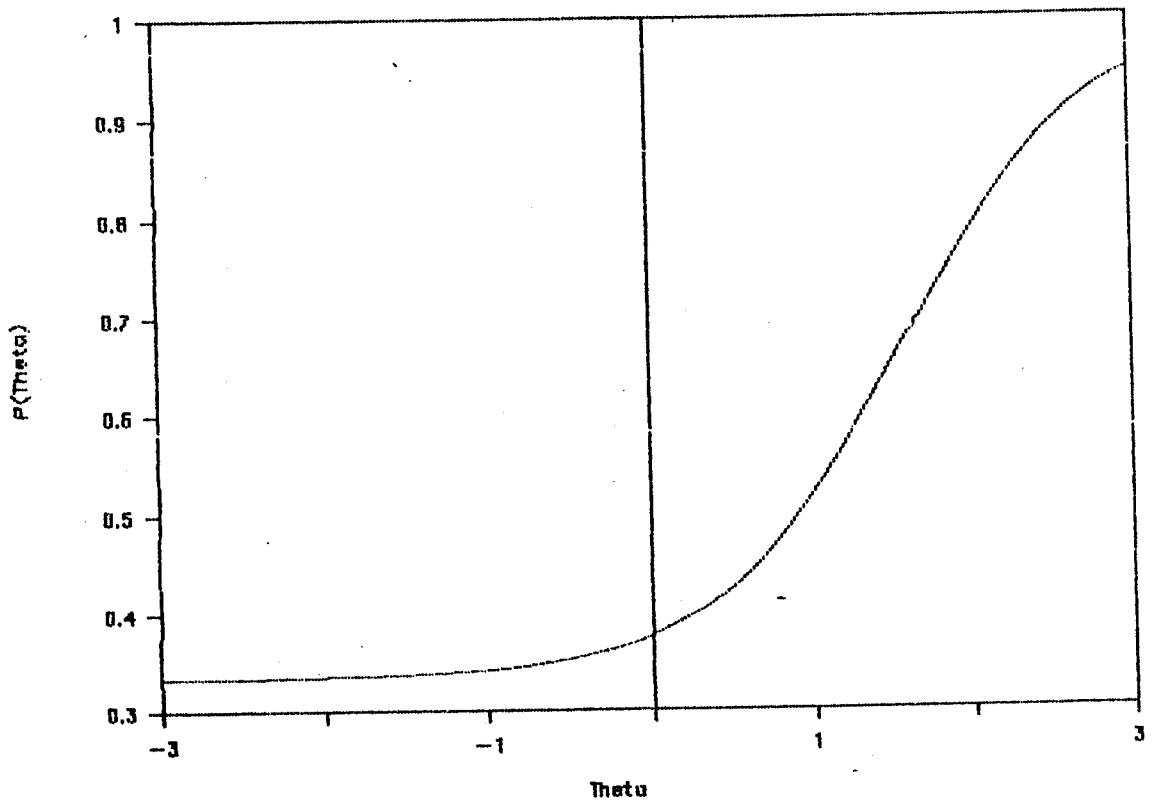
ITEM 31



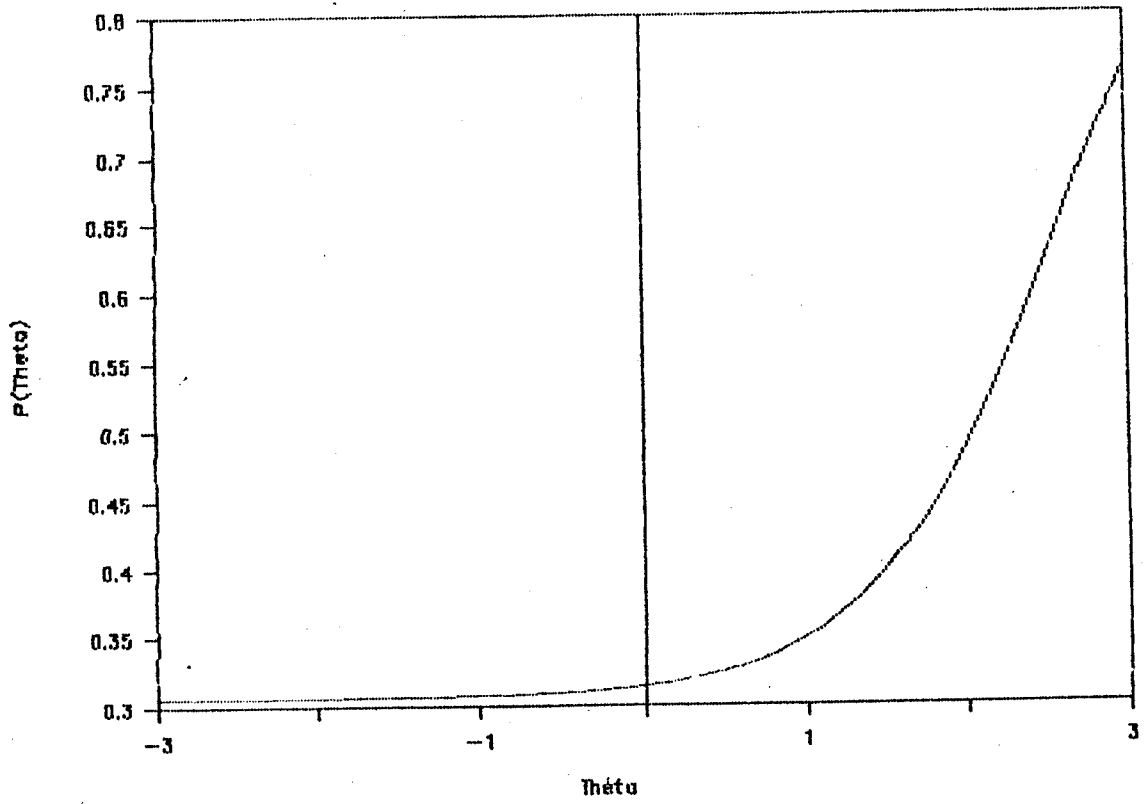
ITEM 32



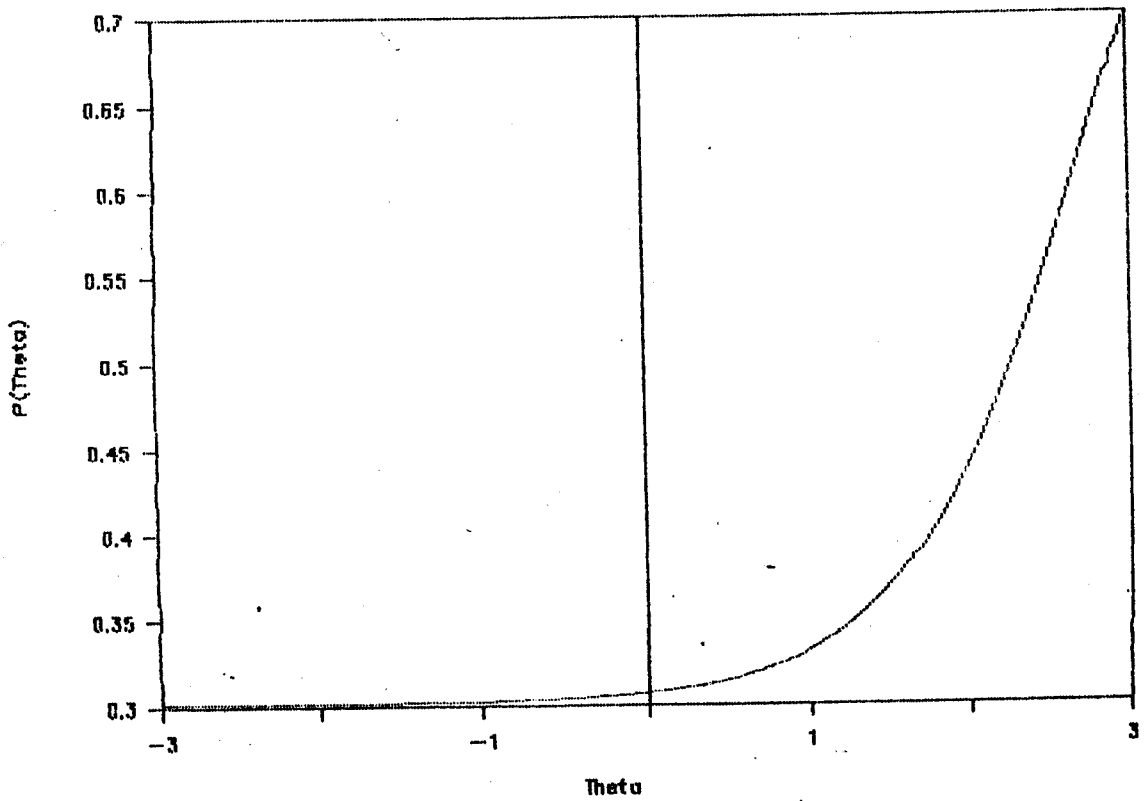
ITEM 33



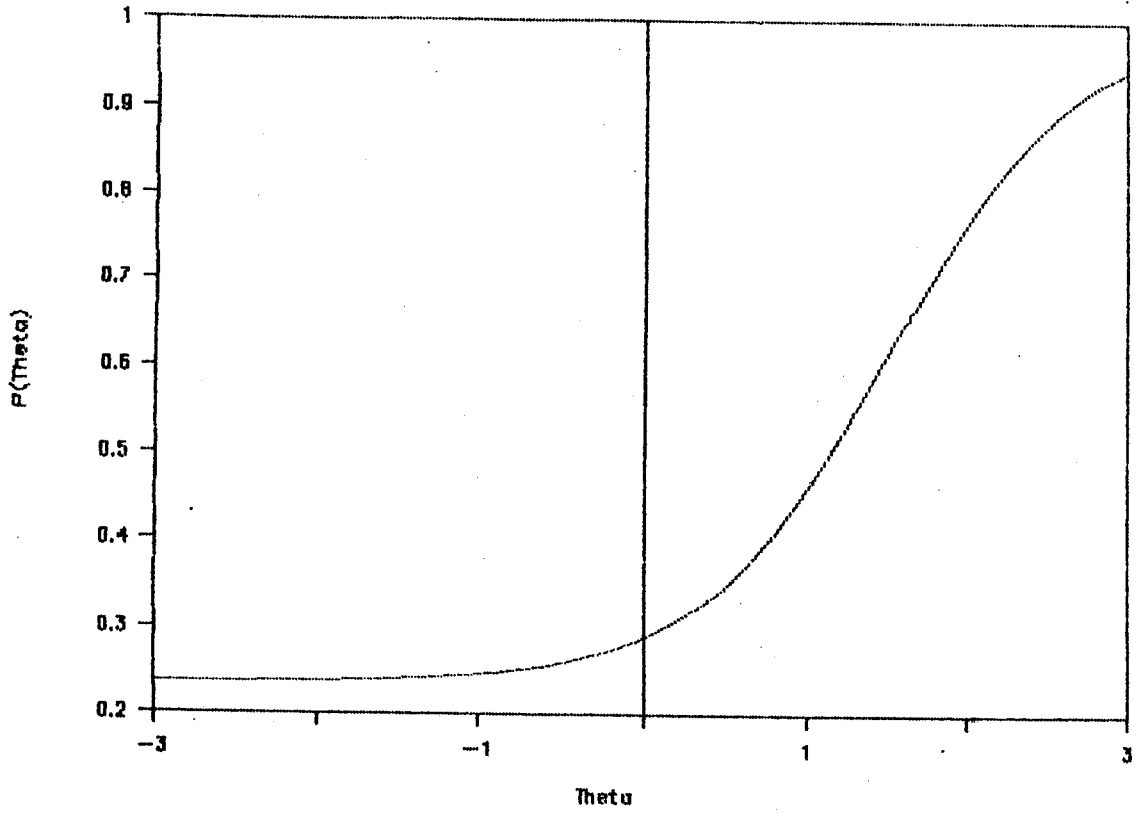
ITEM 34



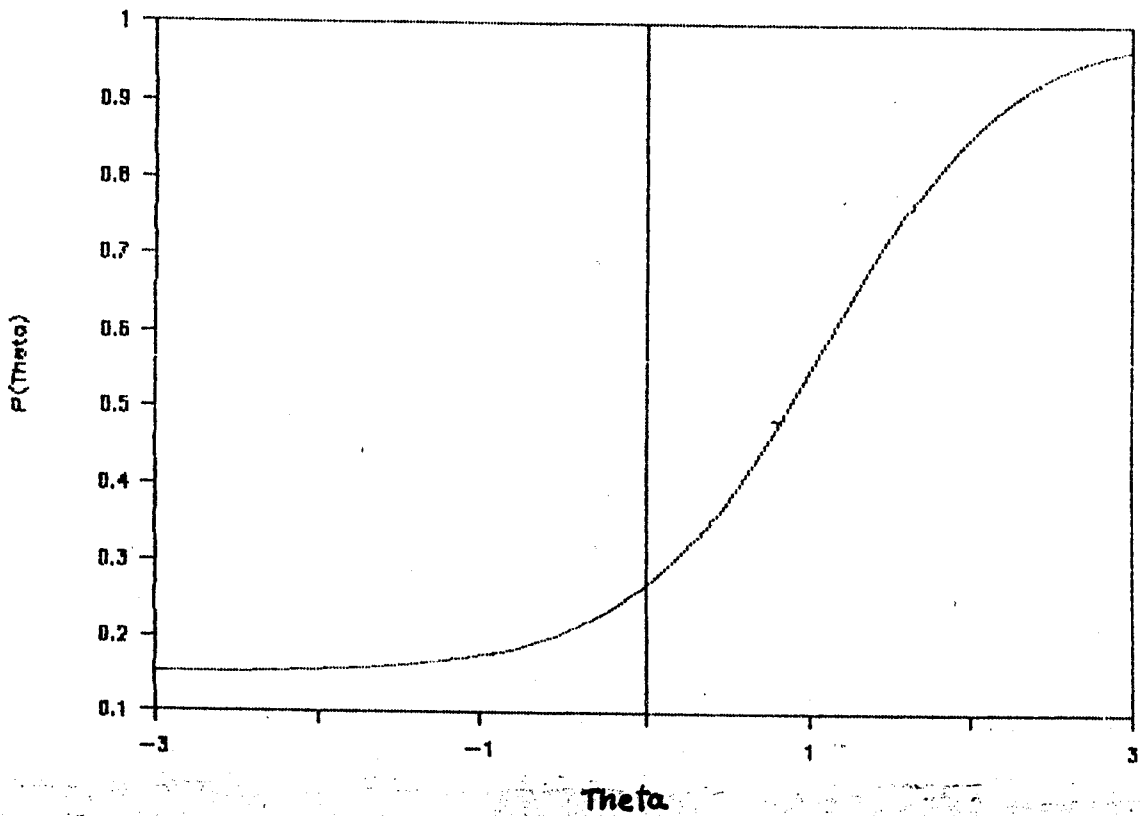
ITEM 35



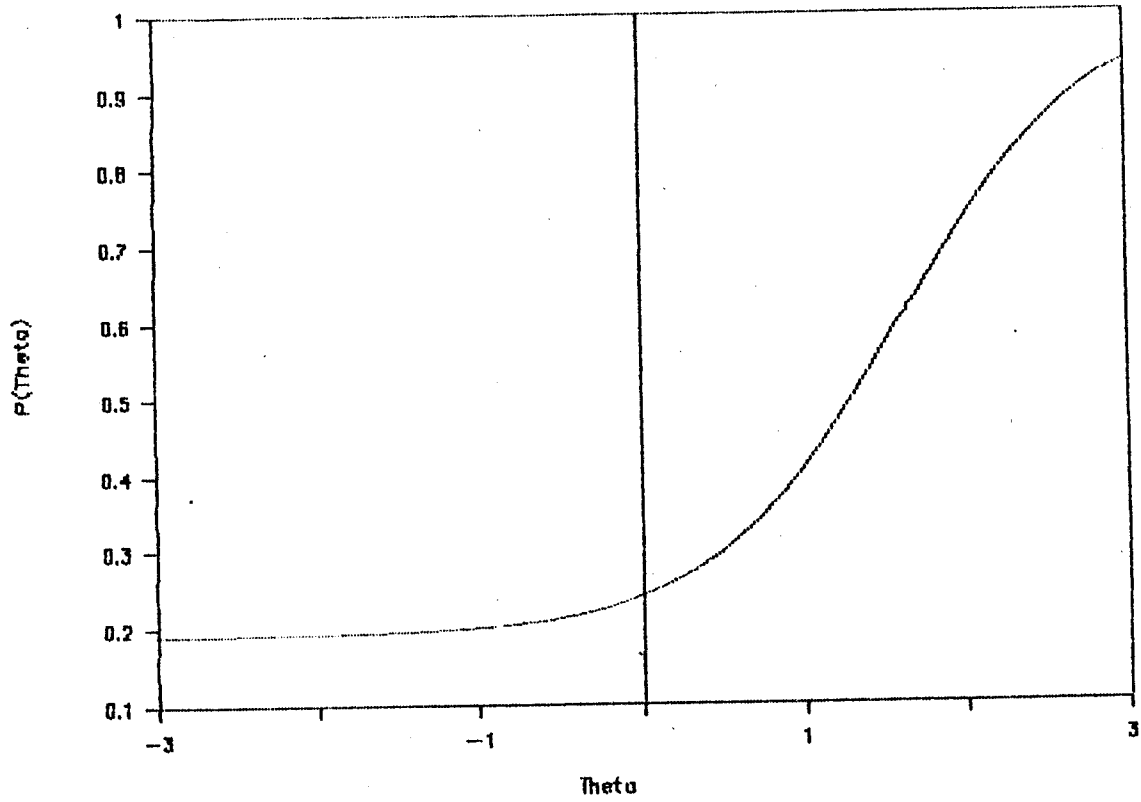
ITEM 36



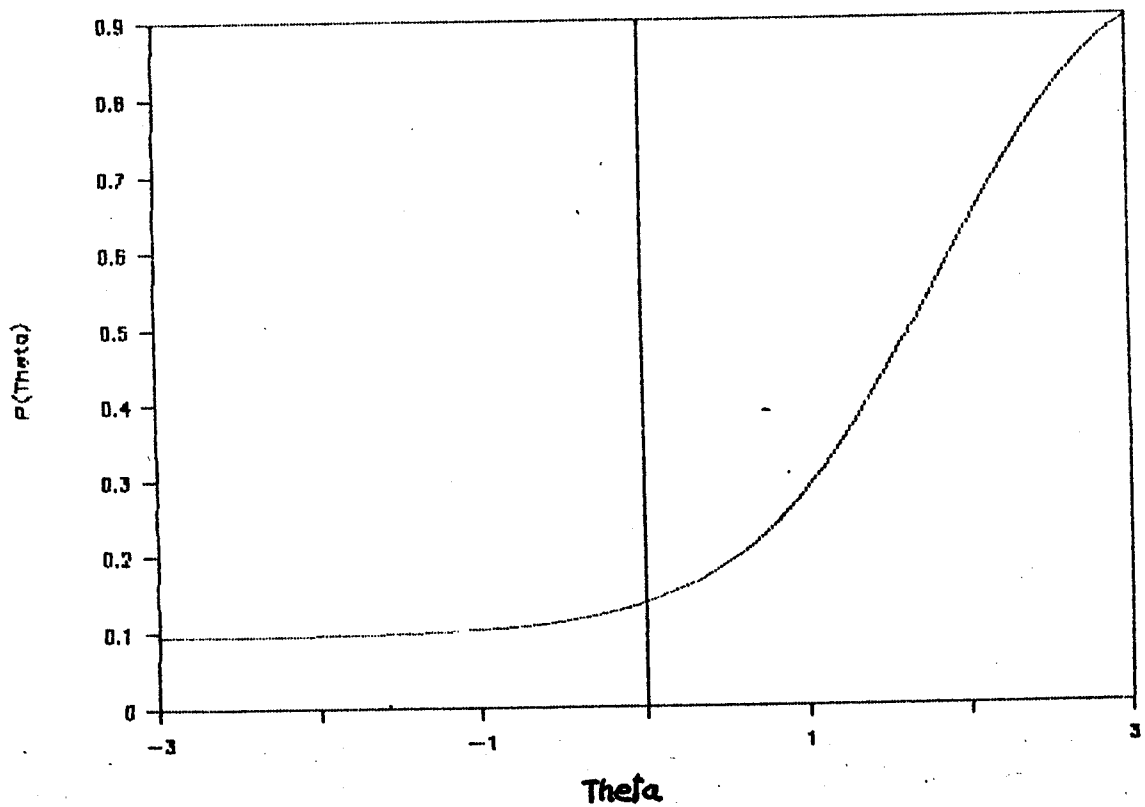
ITEM 37



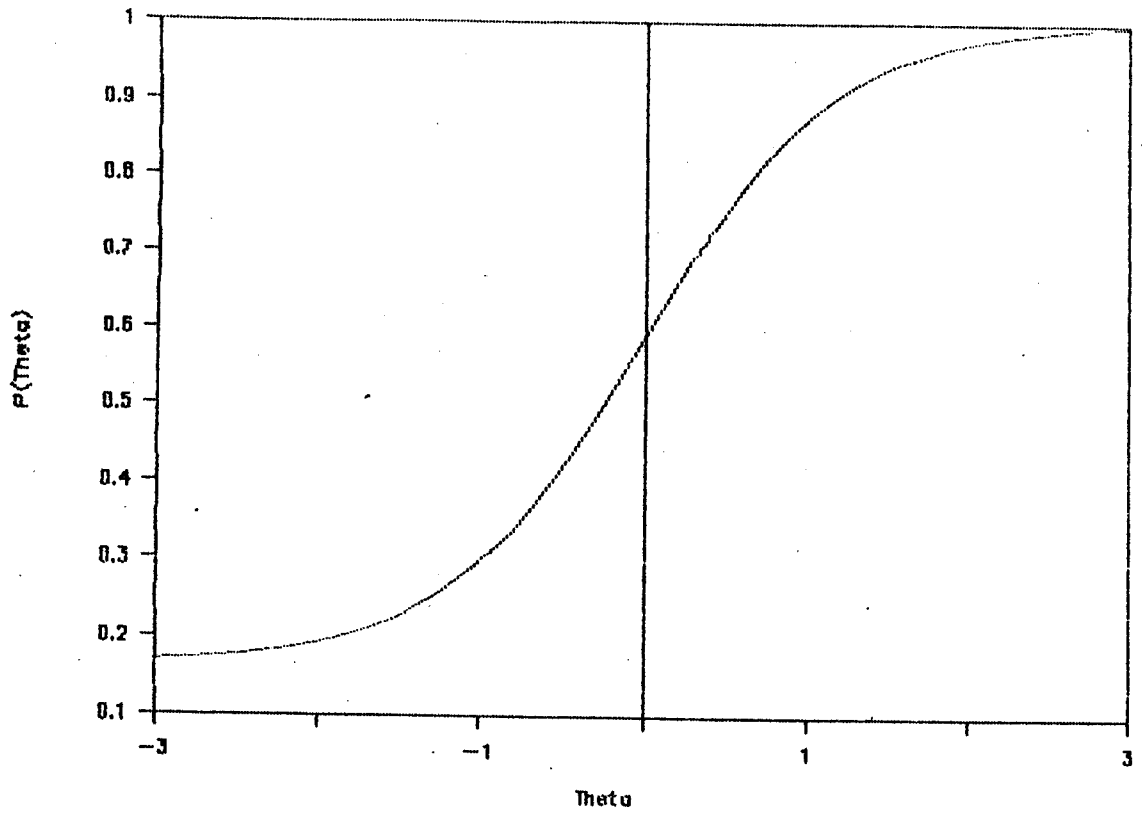
ITEM 38



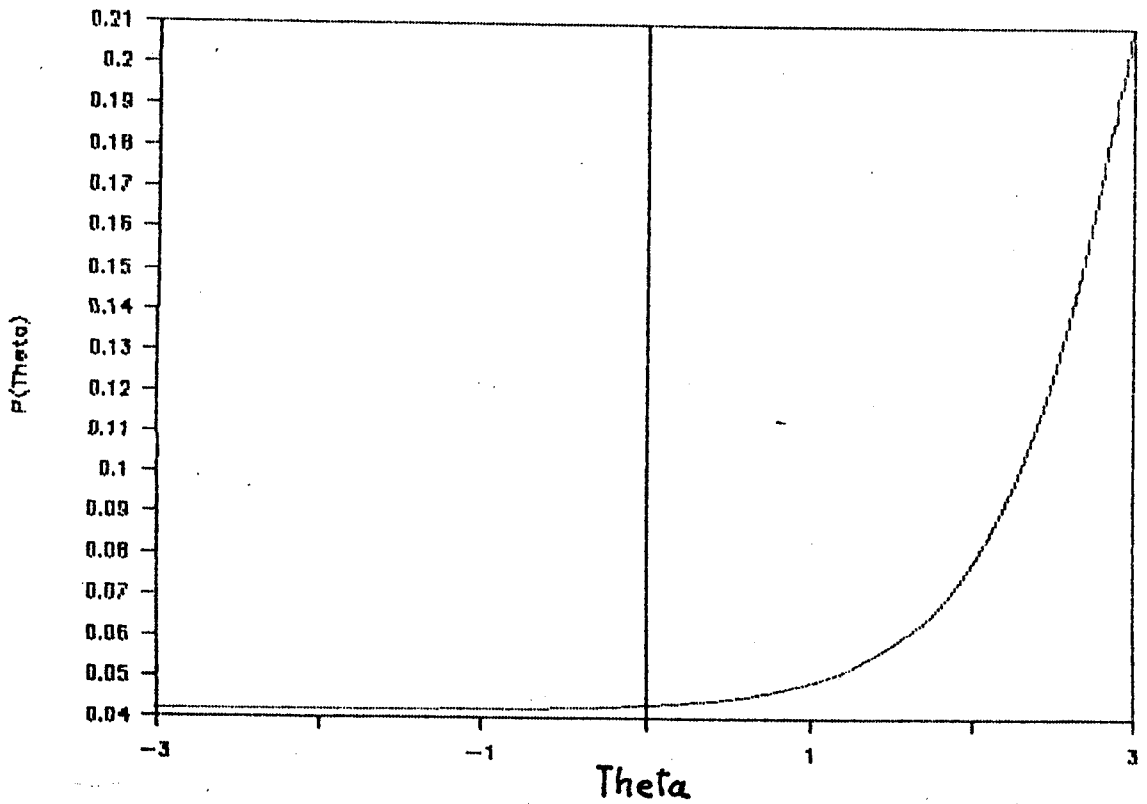
ITEM 39



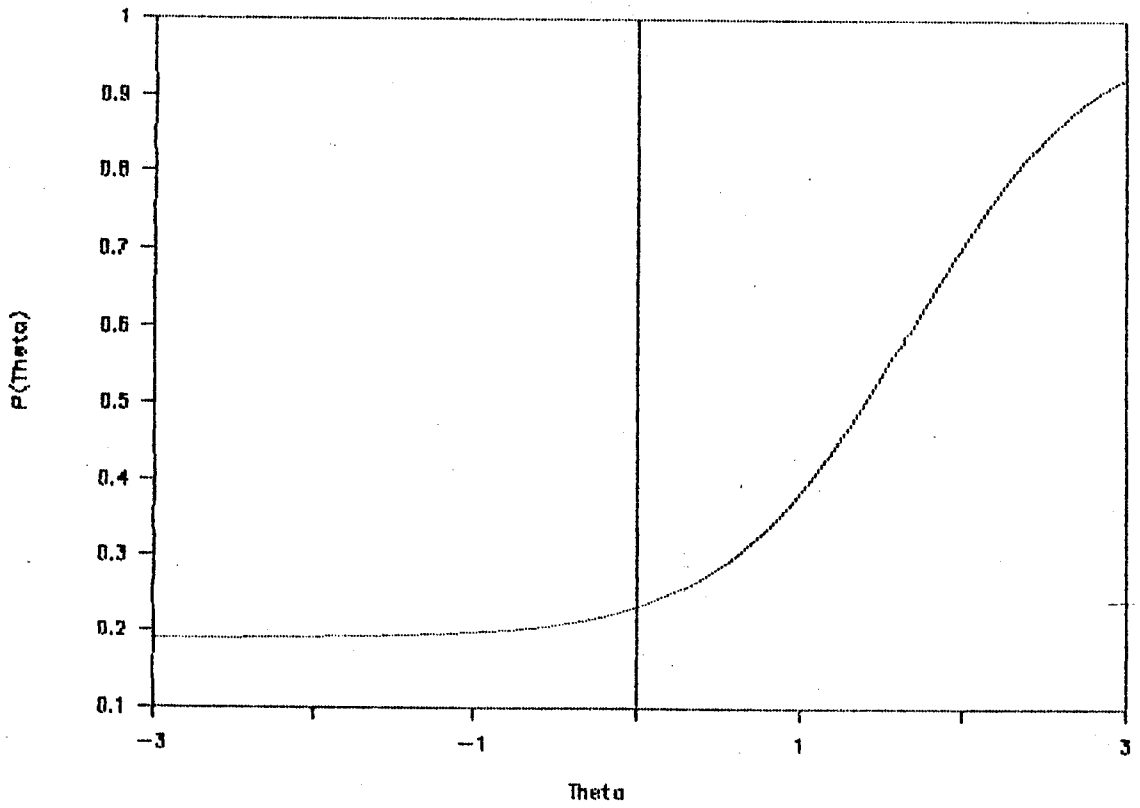
ITEM 40



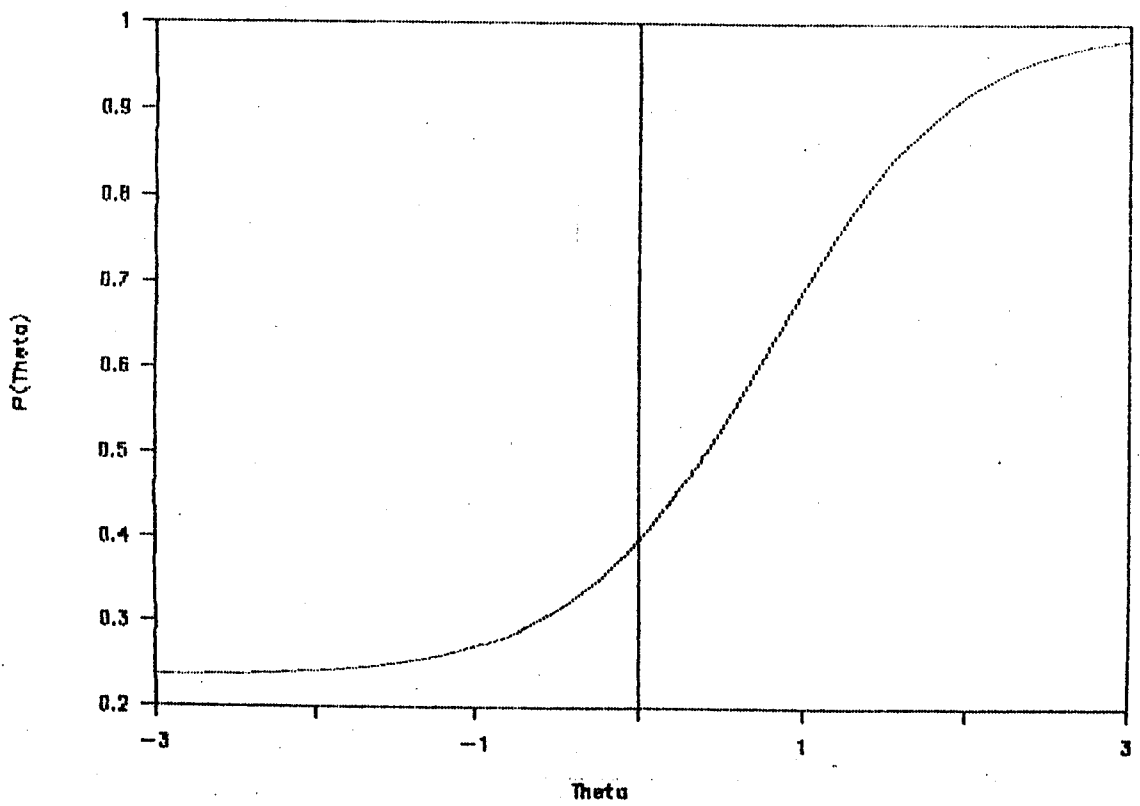
ITEM 41



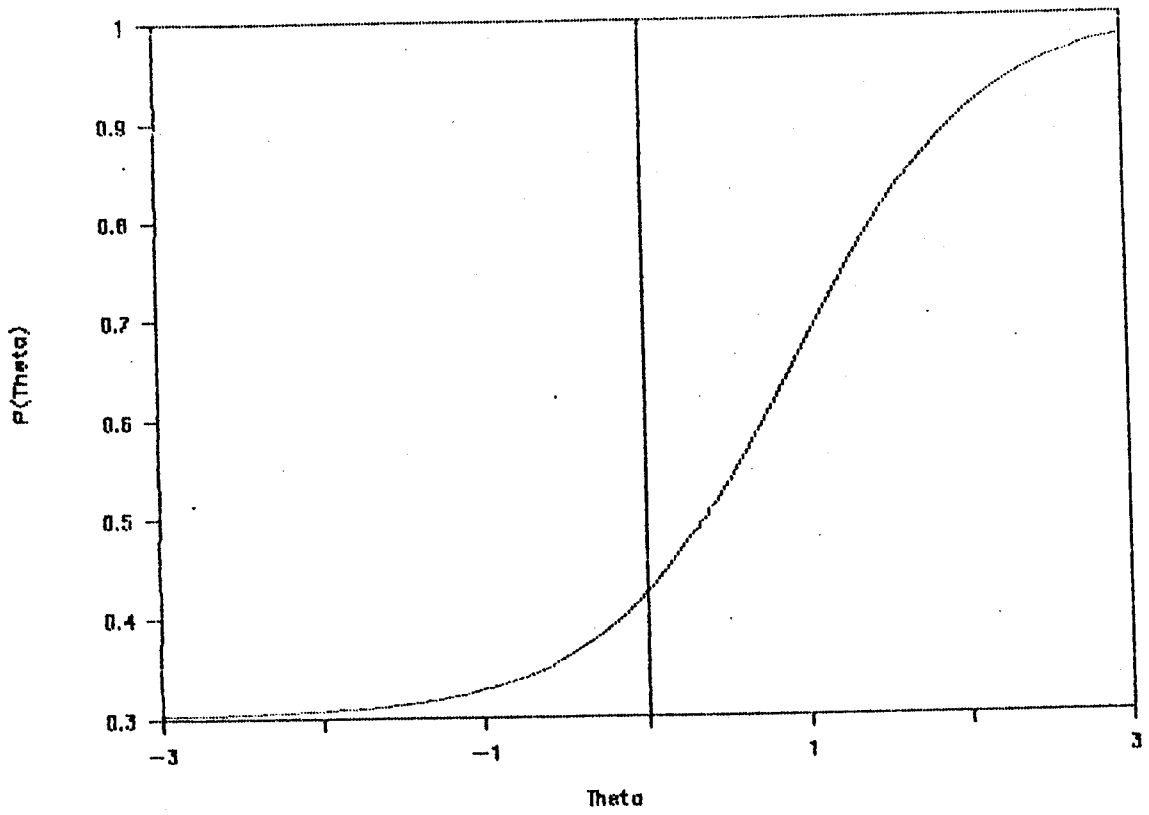
ITEM 42



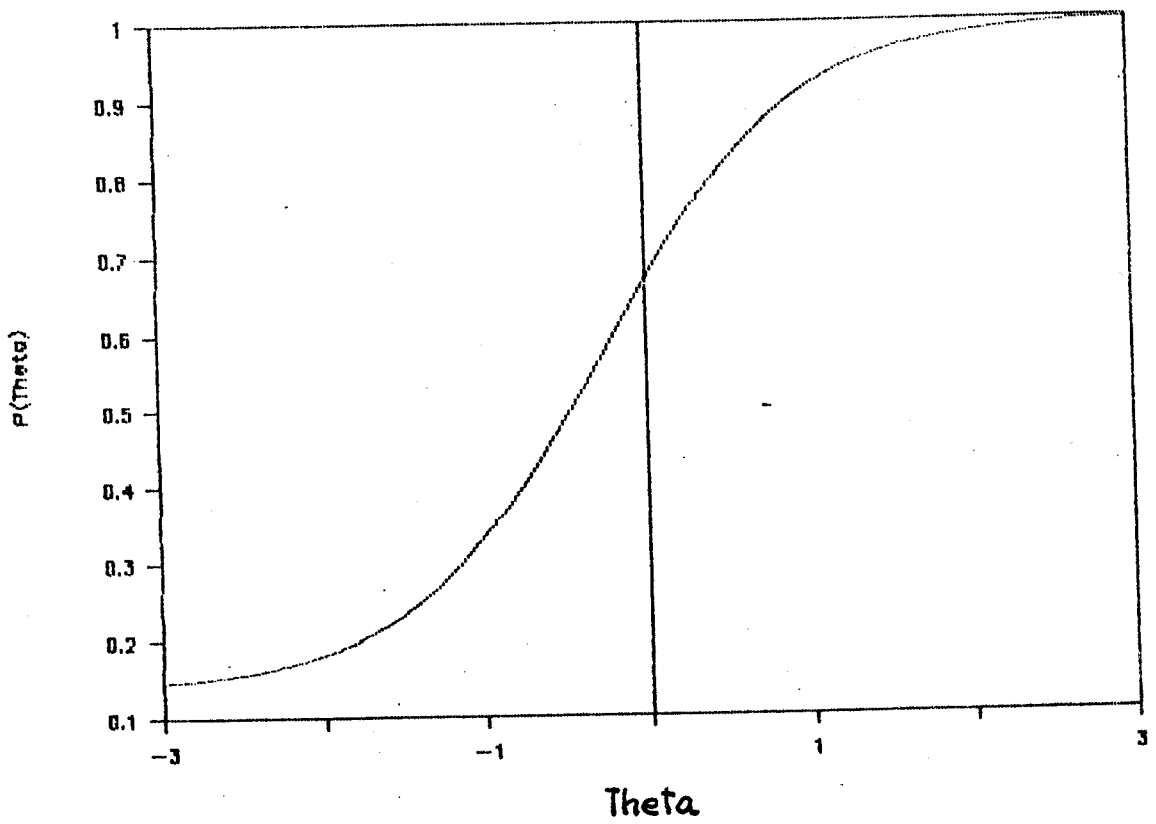
ITEM 43



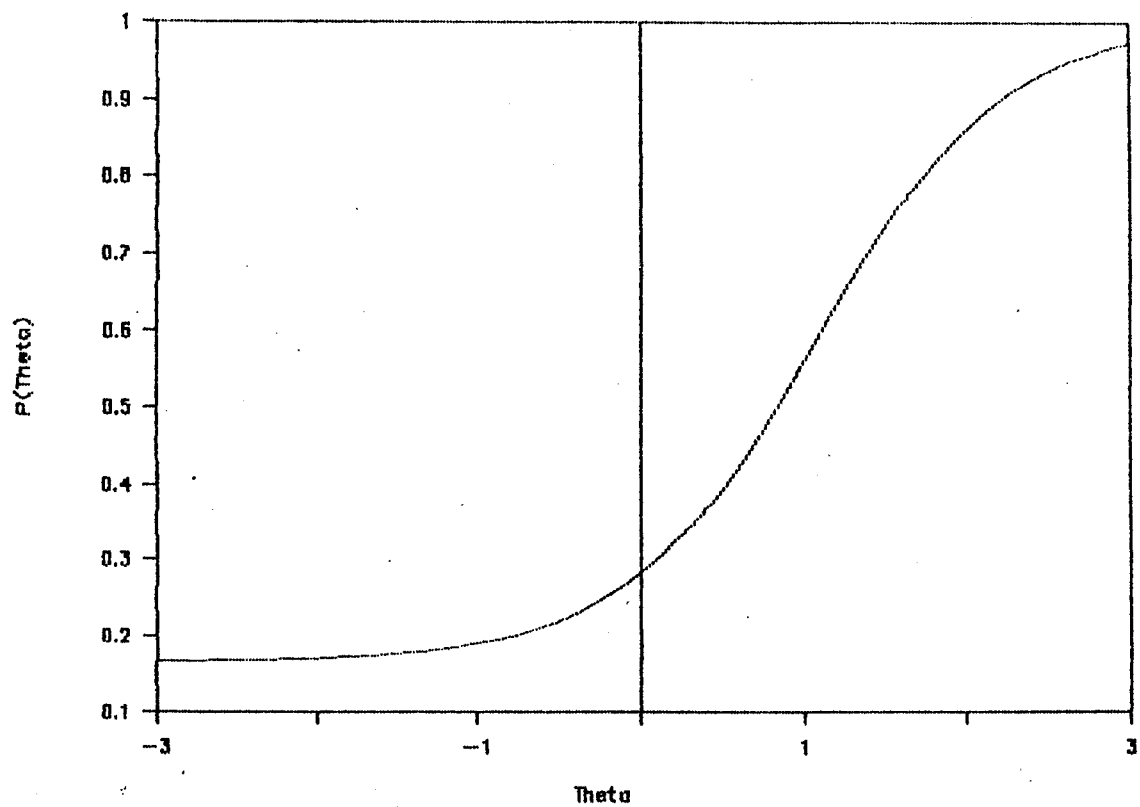
ITEM 44



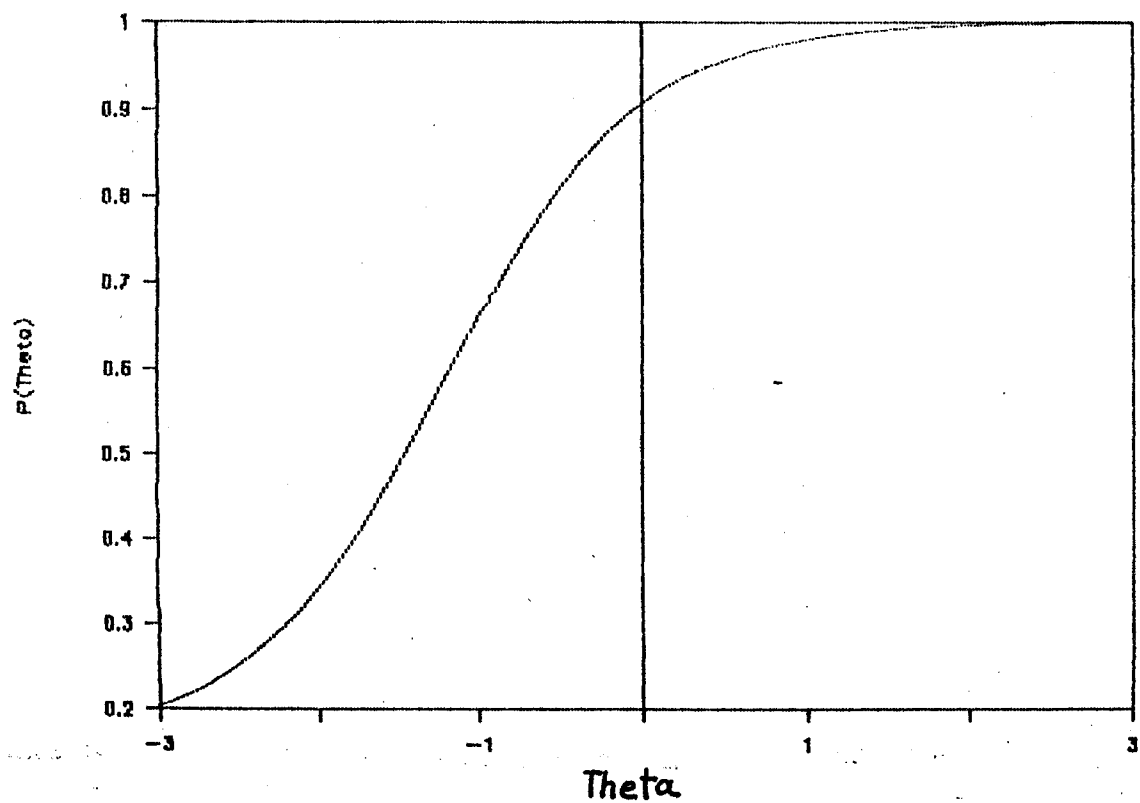
ITEM 45



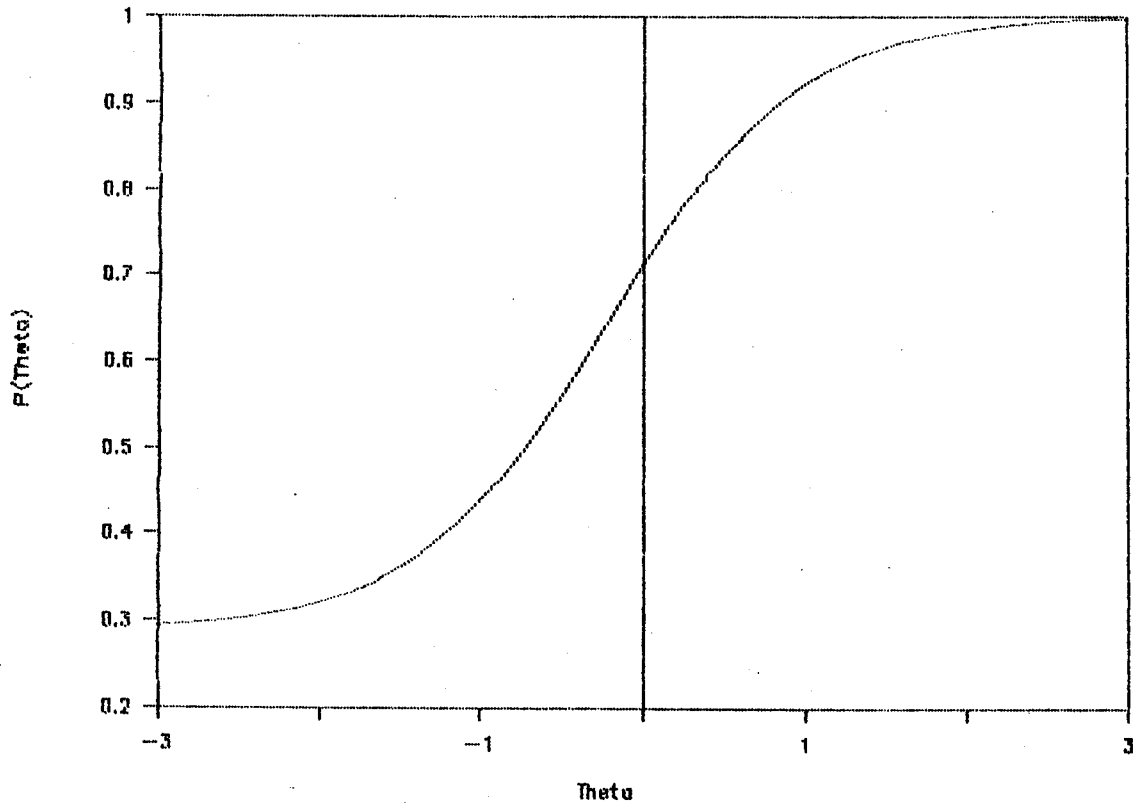
ITEM 46



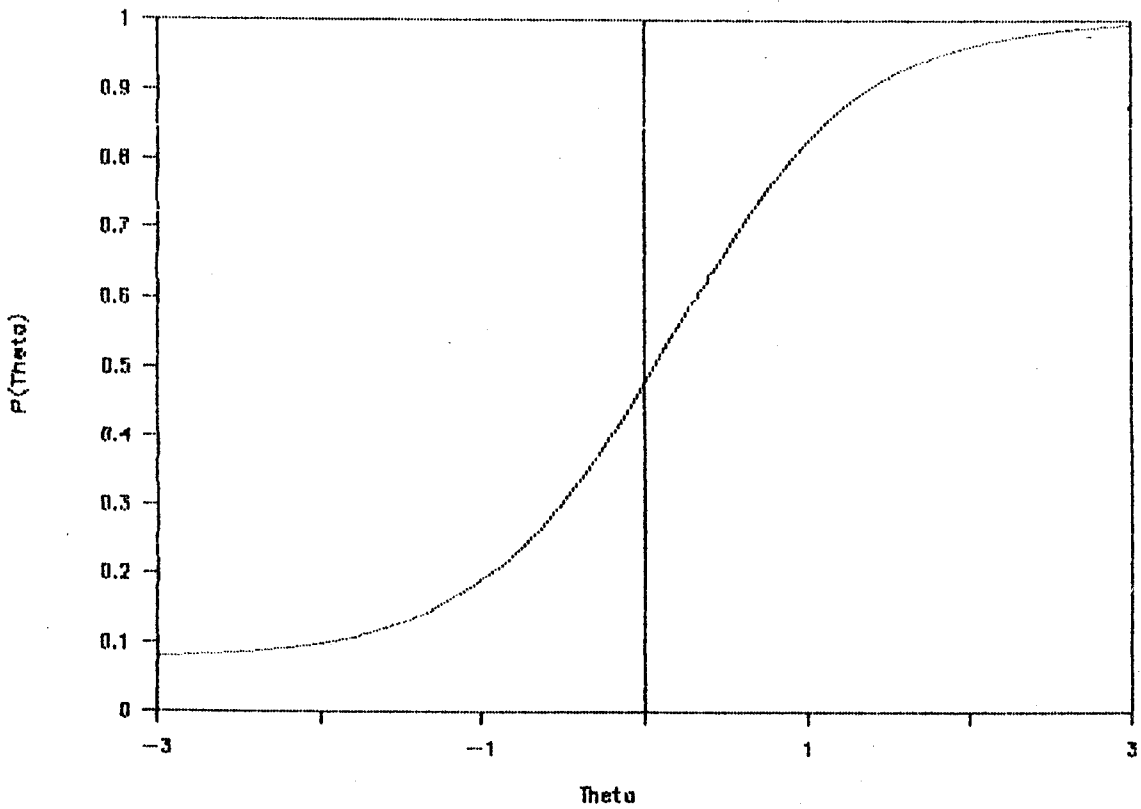
ITEM 47



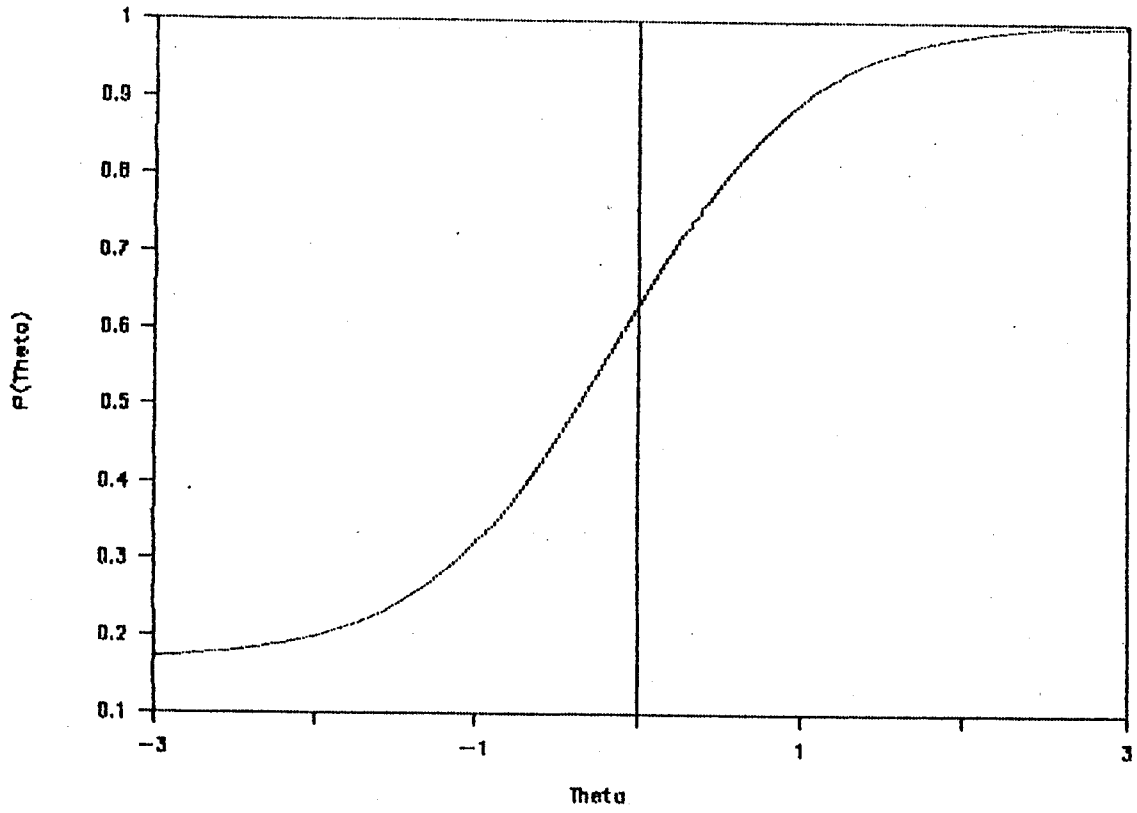
ITEM 48



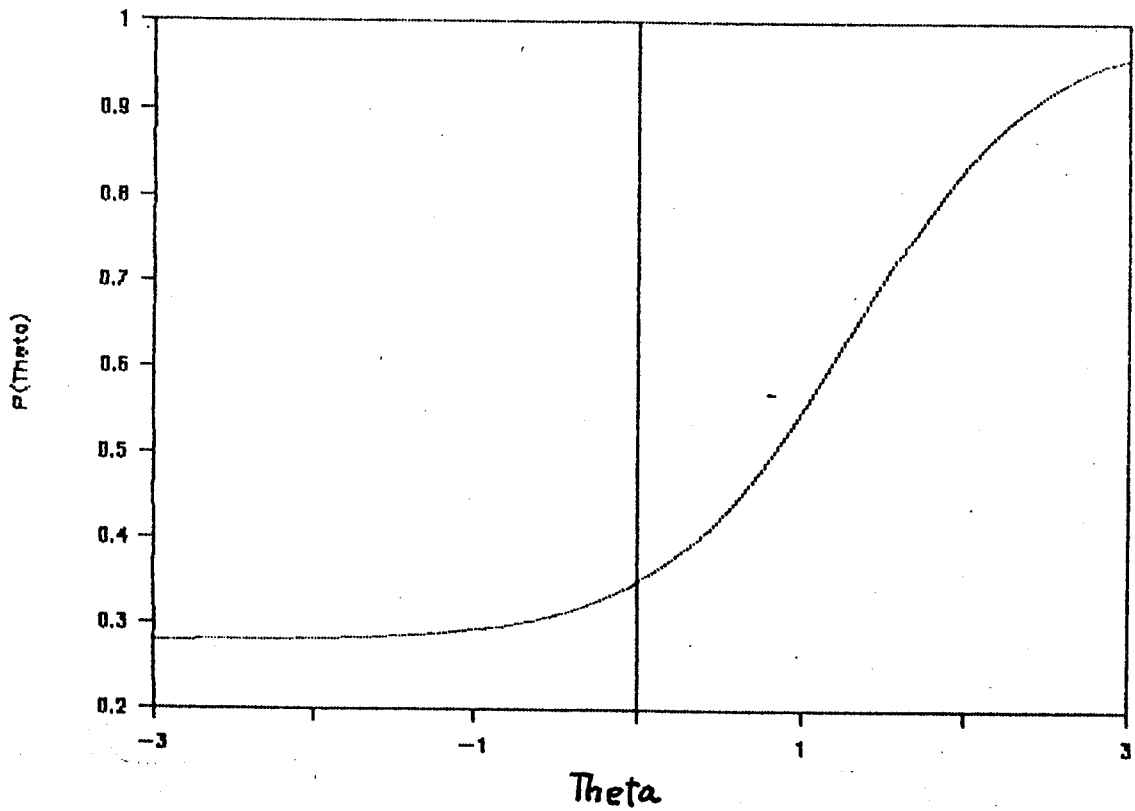
ITEM 49



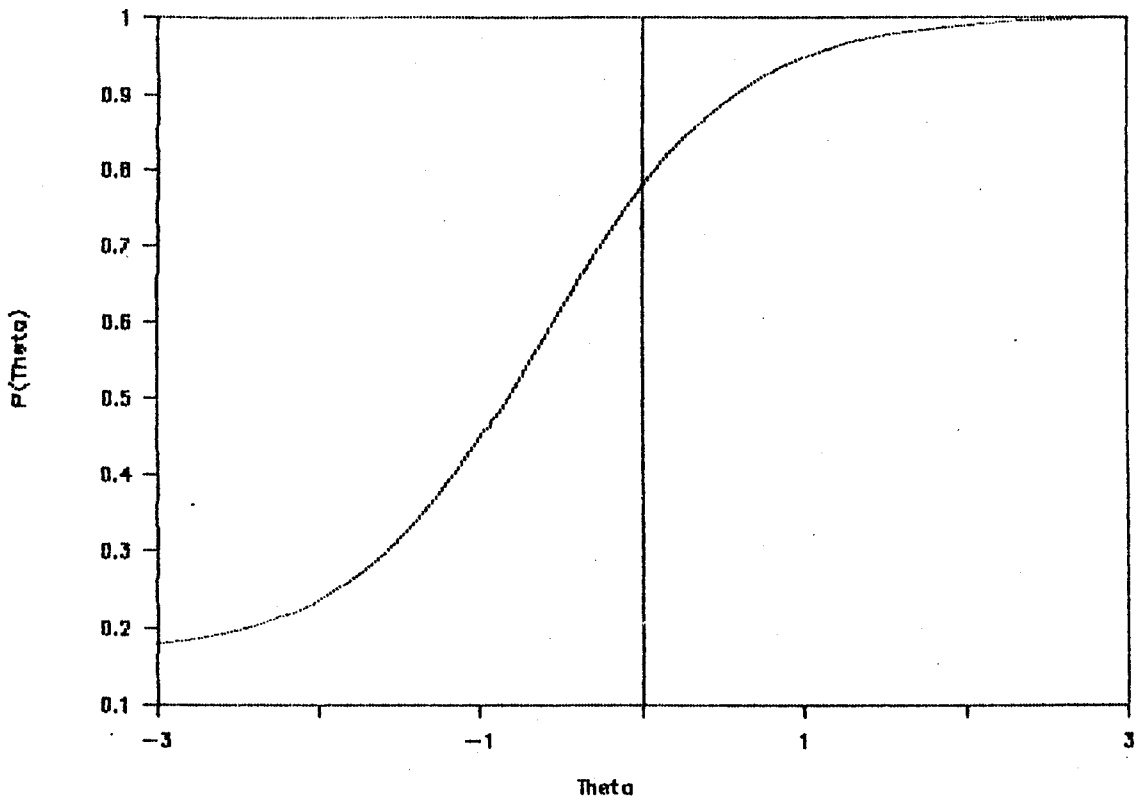
ITEM 50



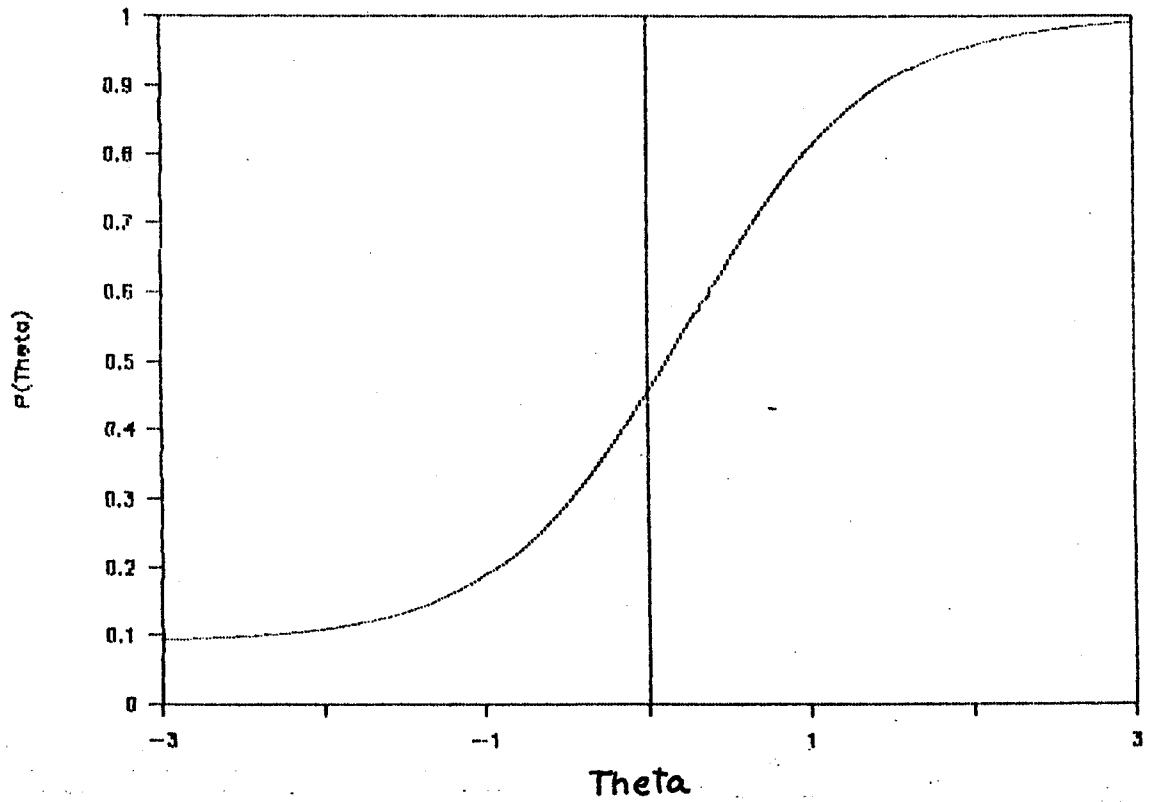
ITEM 51



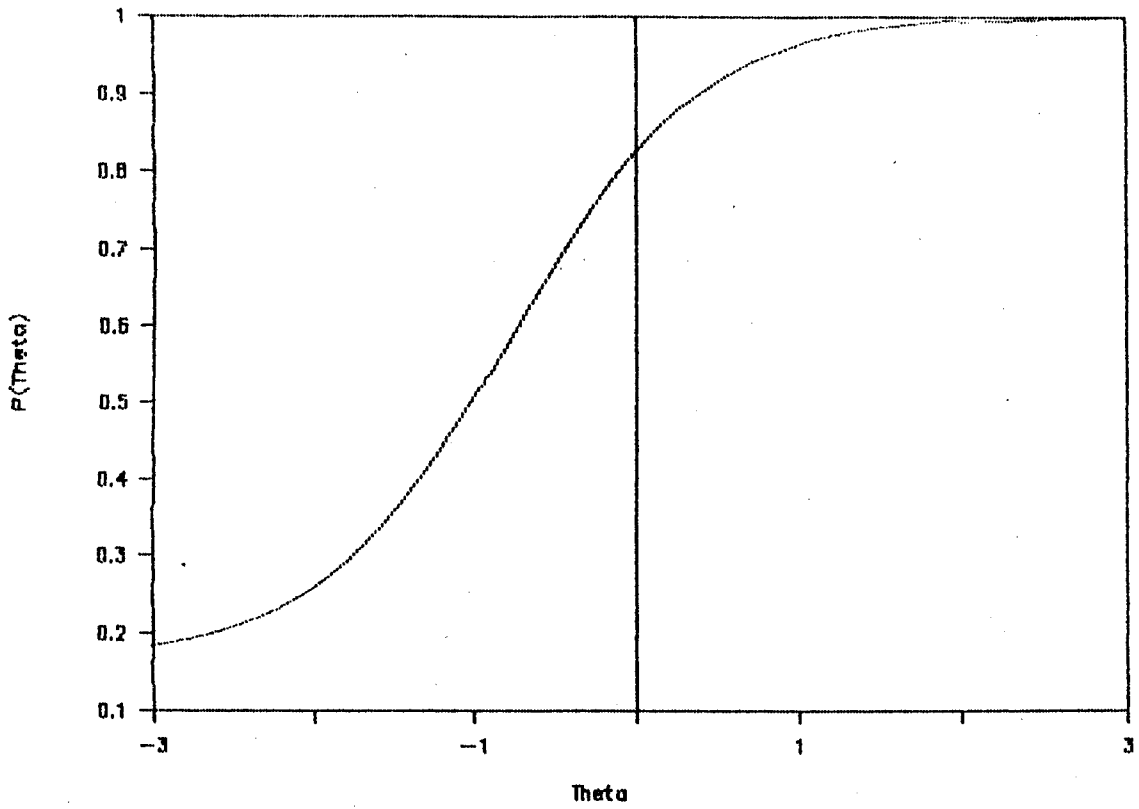
ITEM 52



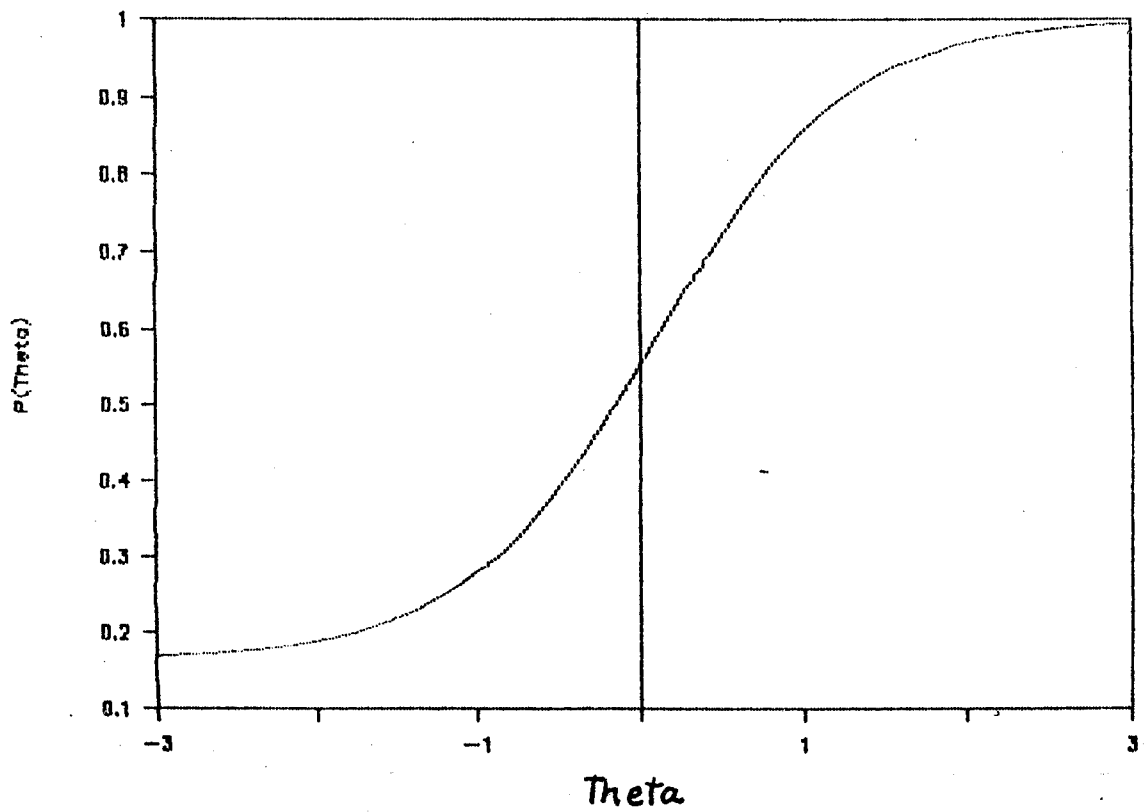
ITEM 53



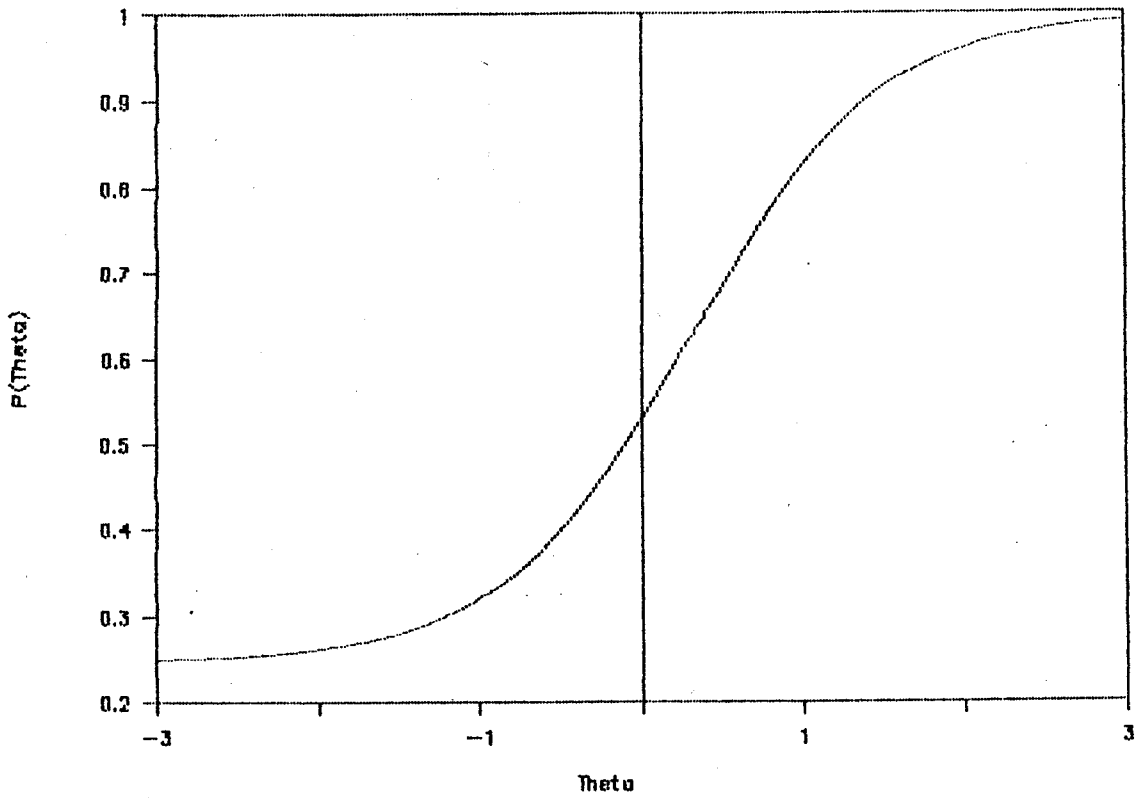
ITEM 54



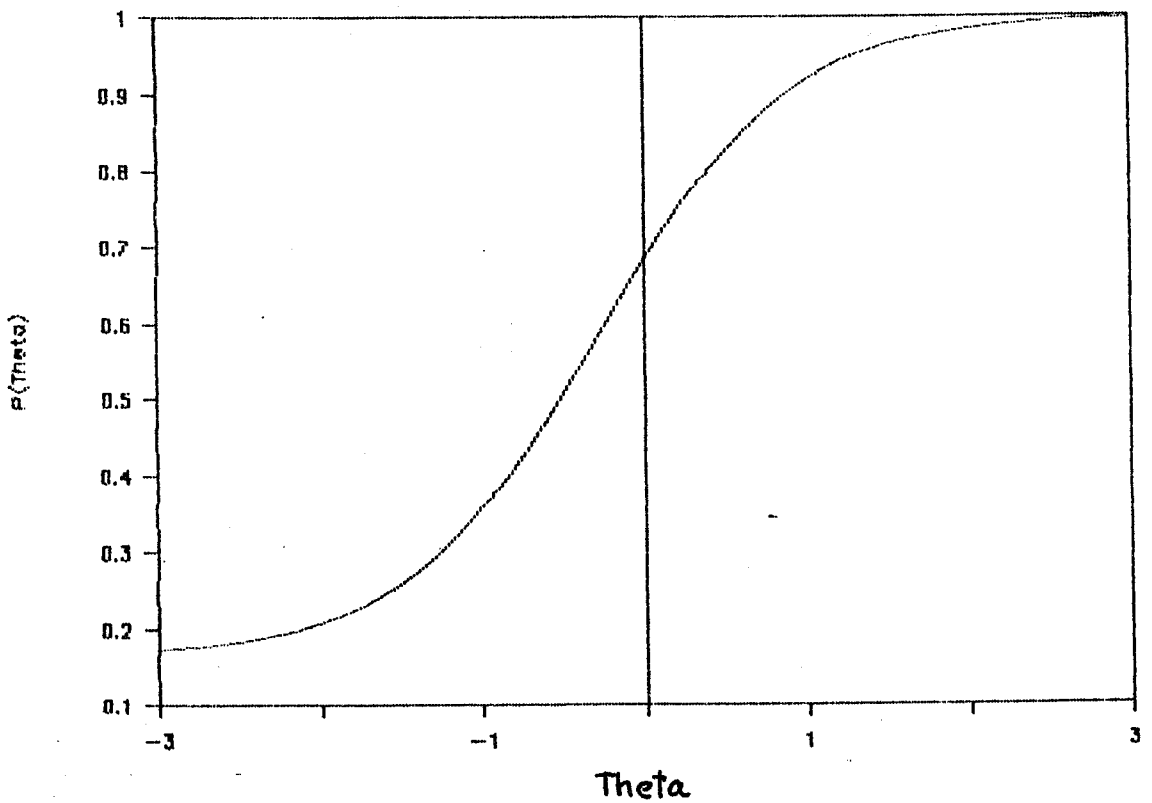
ITEM 55



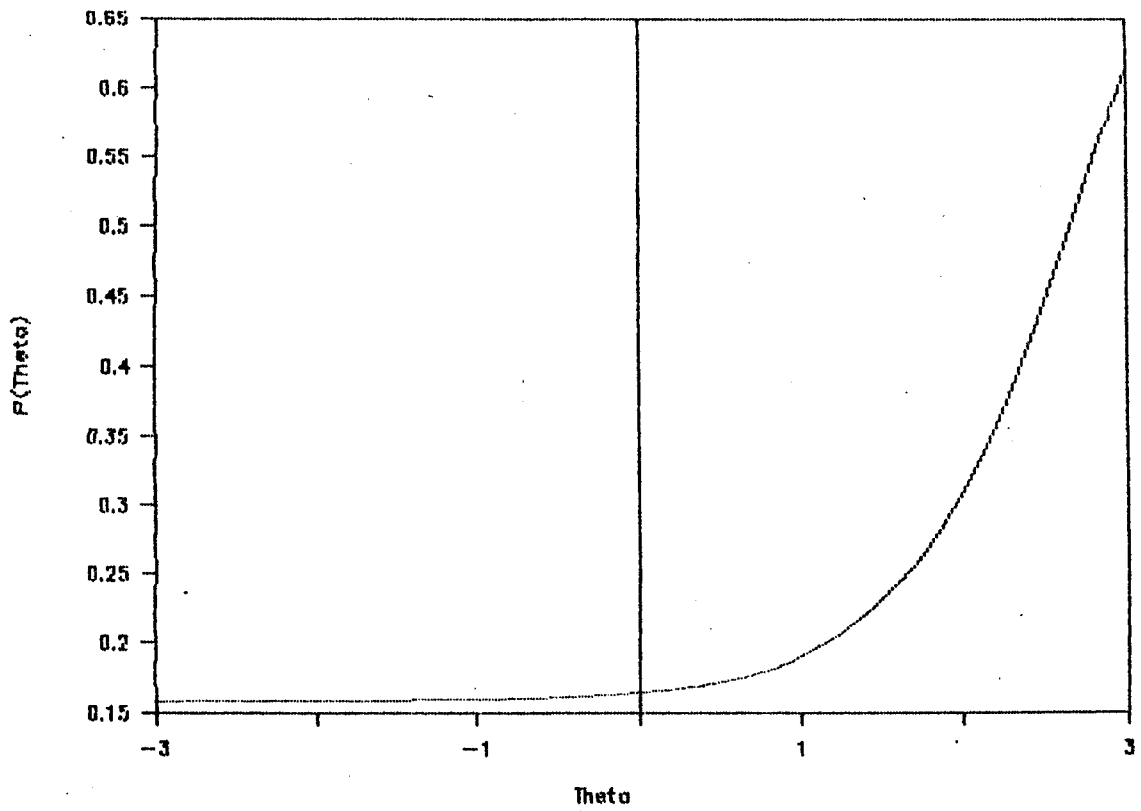
ITEM 56



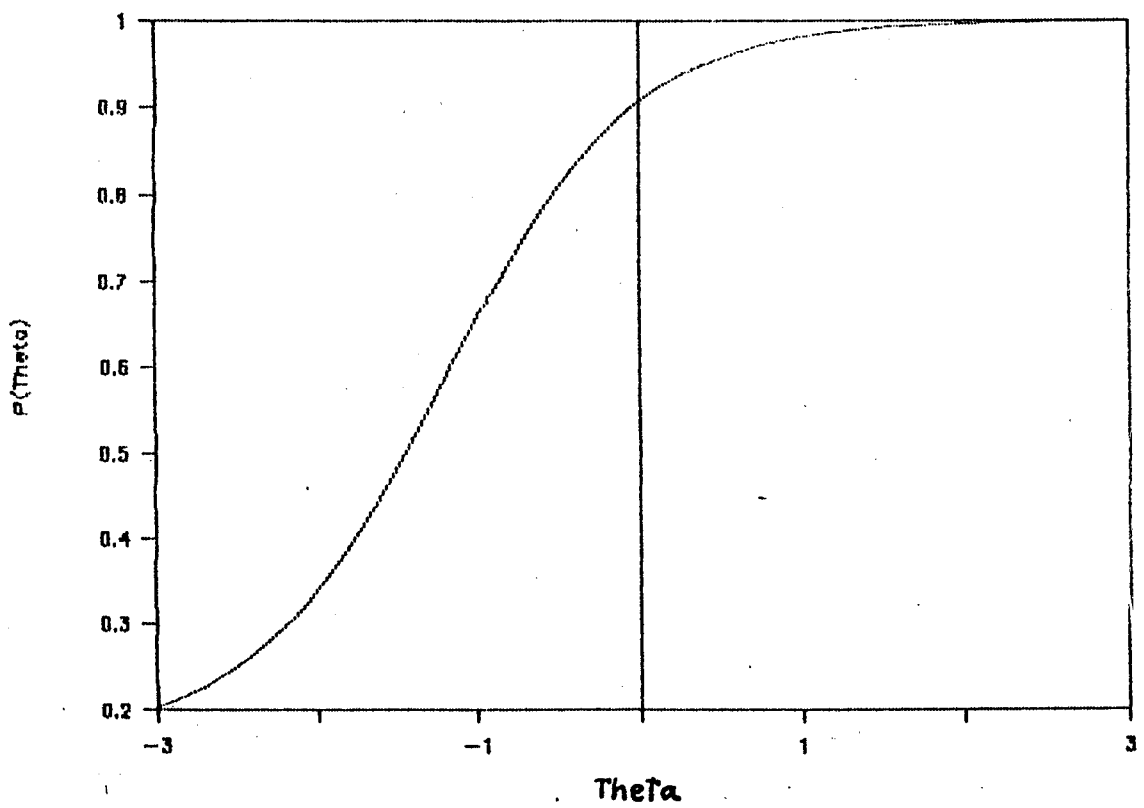
ITEM 57



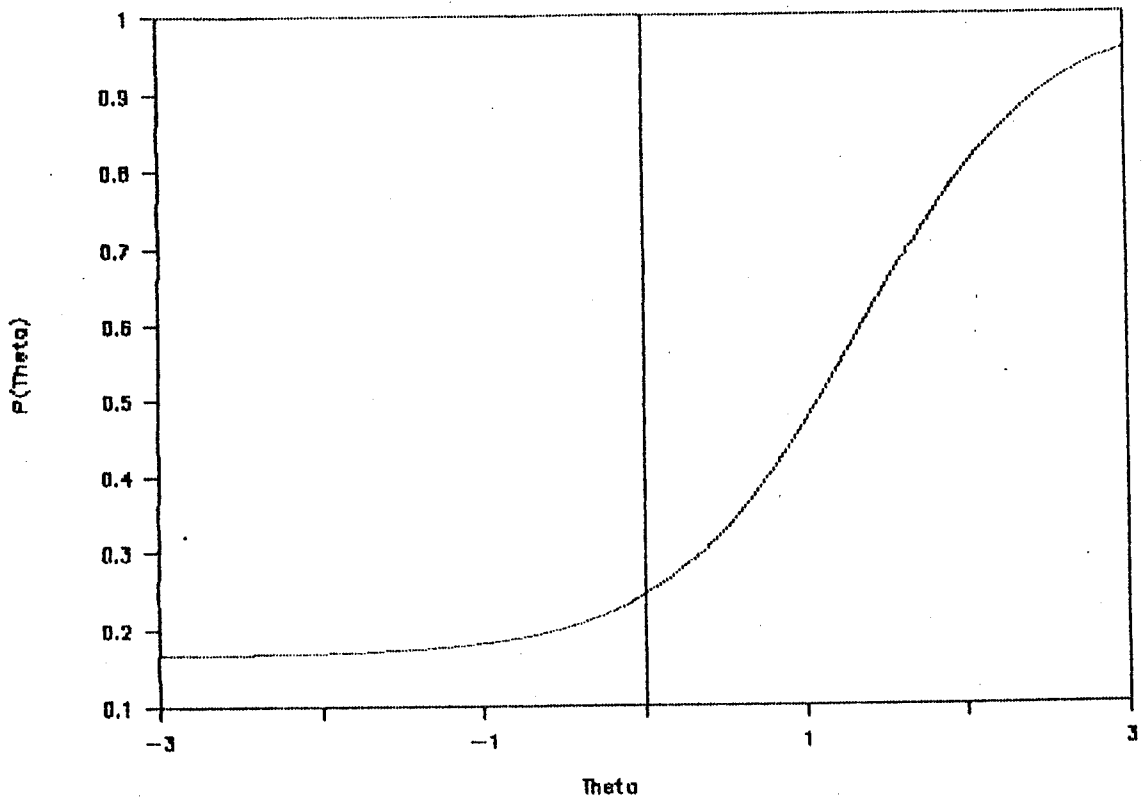
ITEM 58



ITEM 59



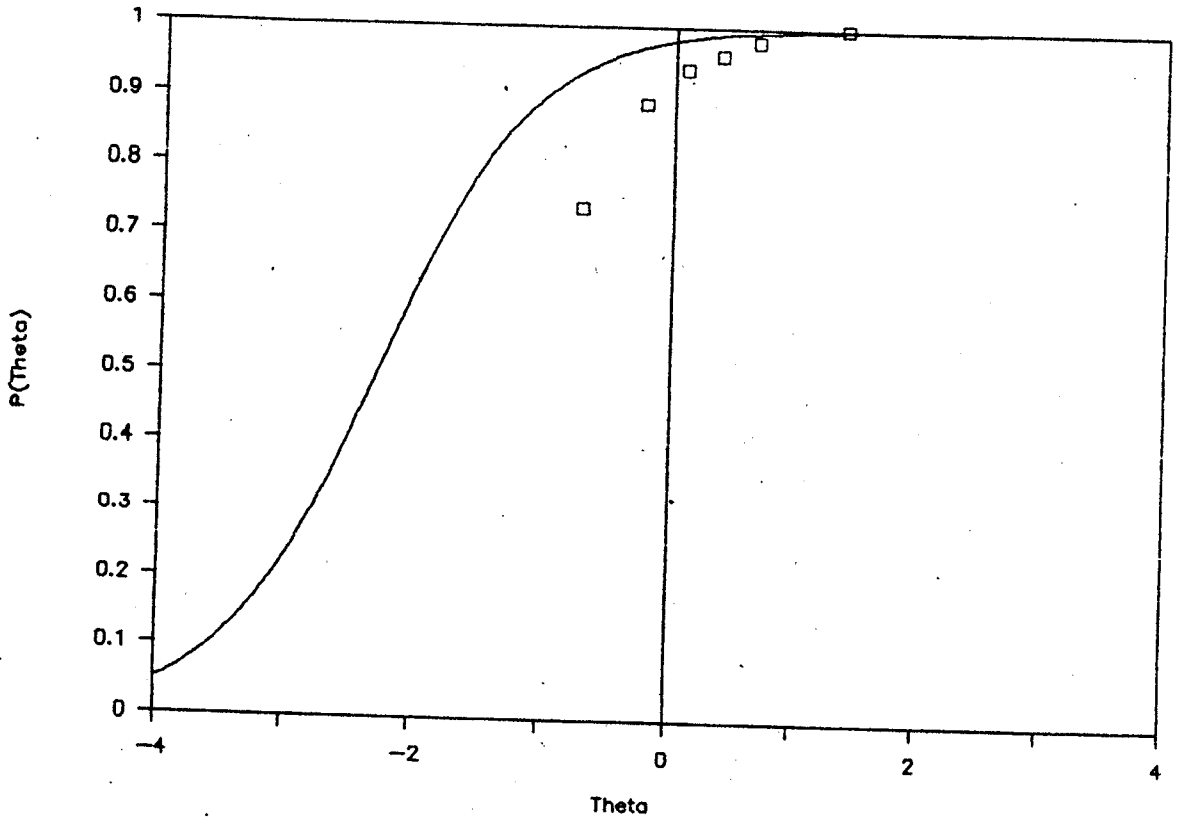
ITEM 60



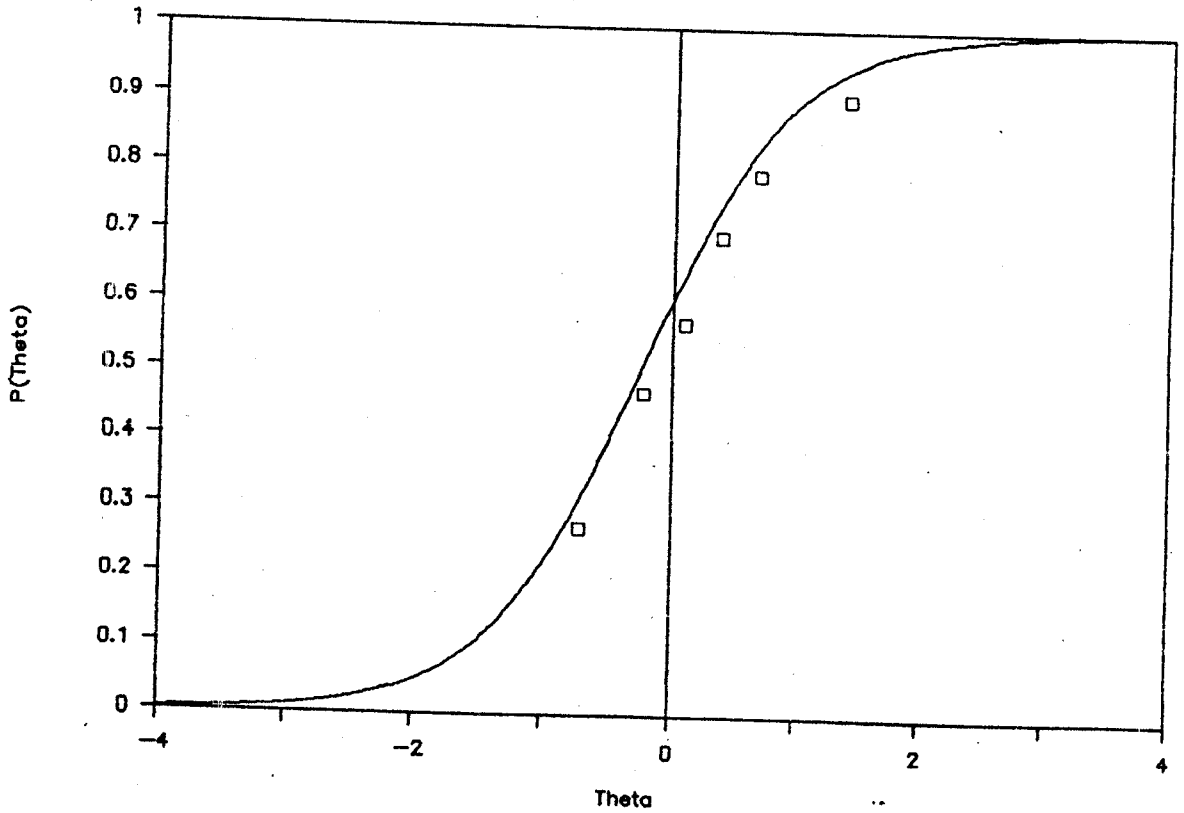
ภาคผนวก ค

ได้แสดงลักษณะของข้อคำถาม 59 ข้อ ในแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน
วิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ที่งบประมาณพารามิเตอร์ภายใต้โมเดล 1PL

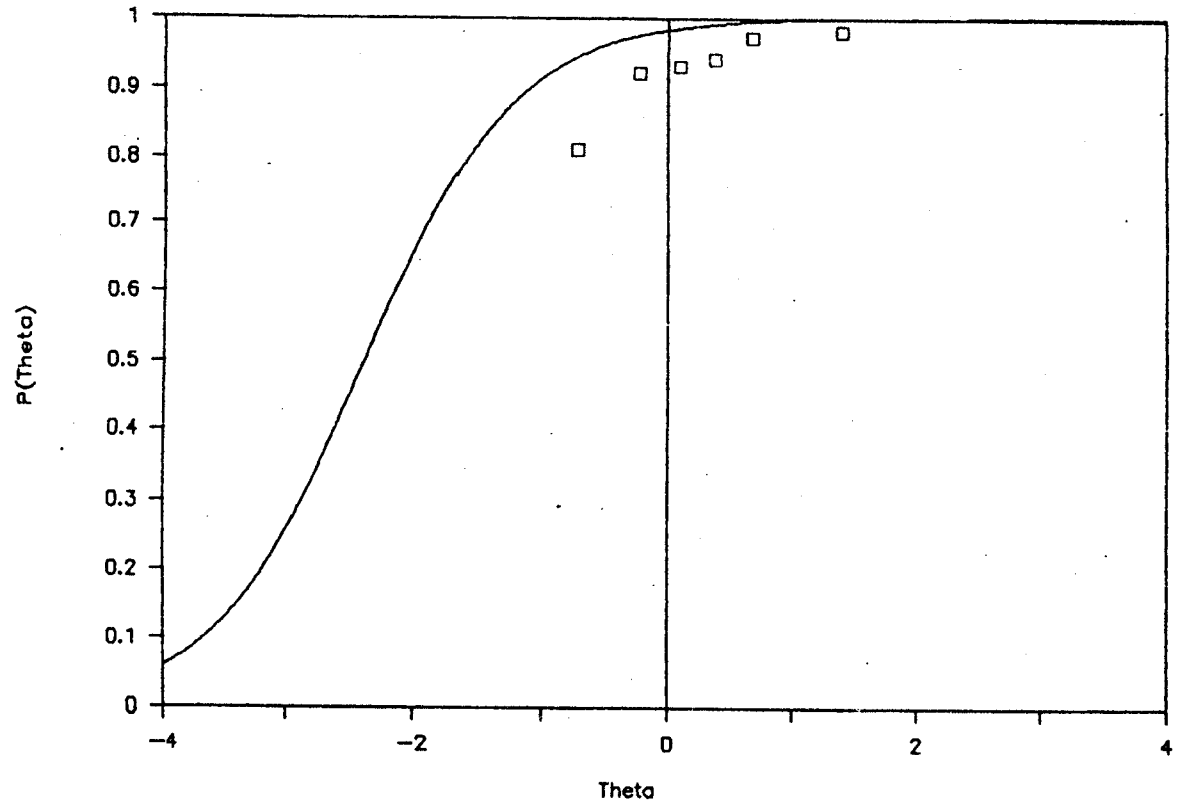
ITEM 01



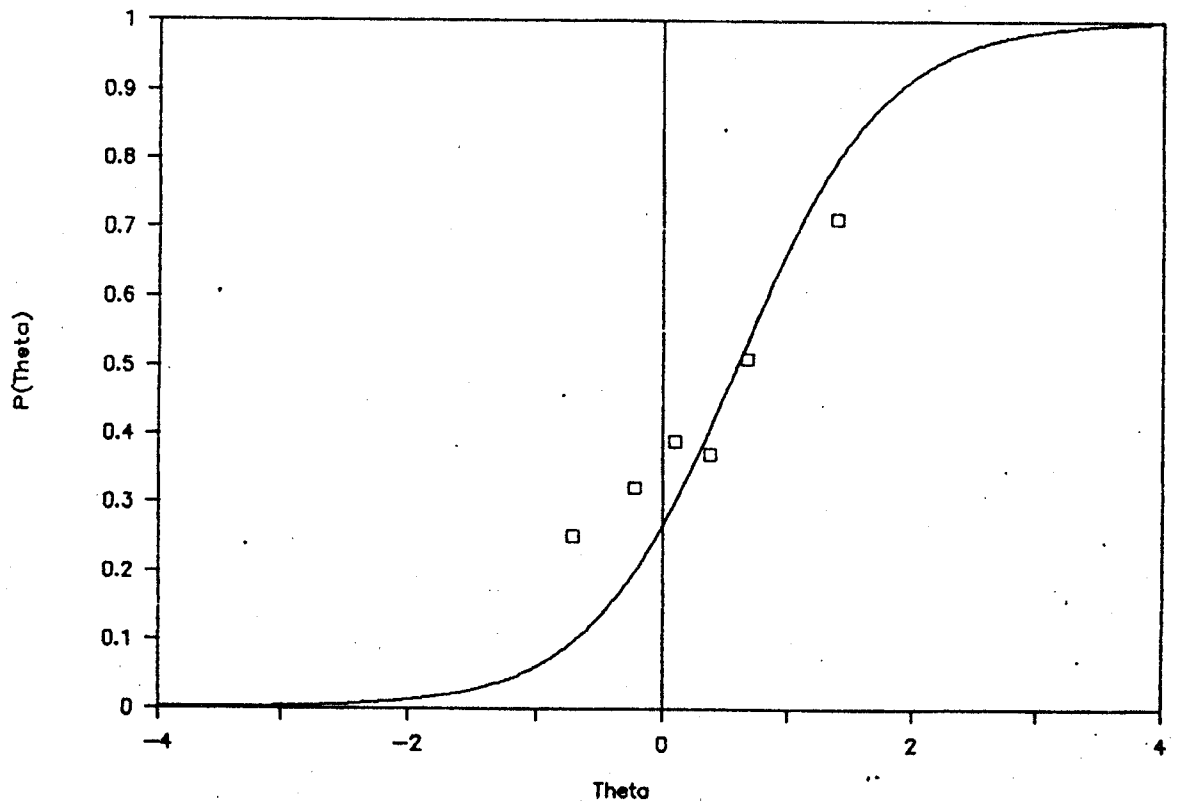
ITEM 02



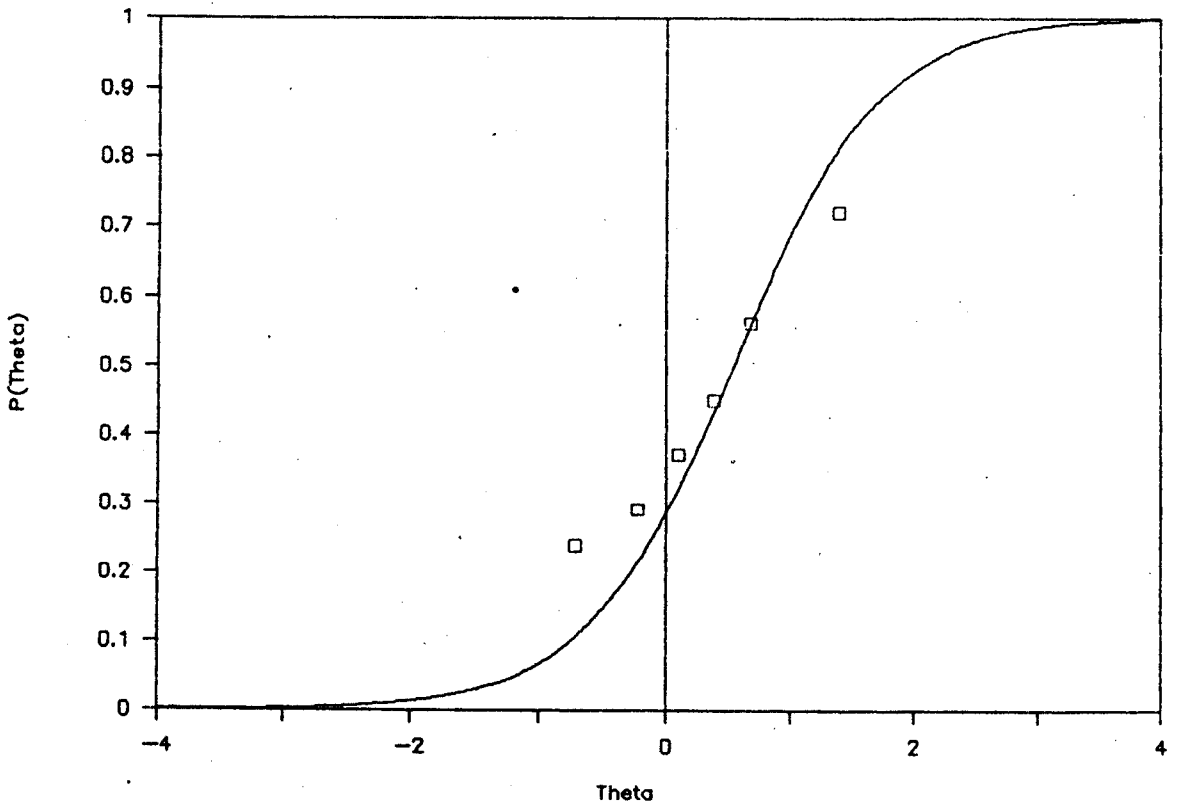
ITEM 03



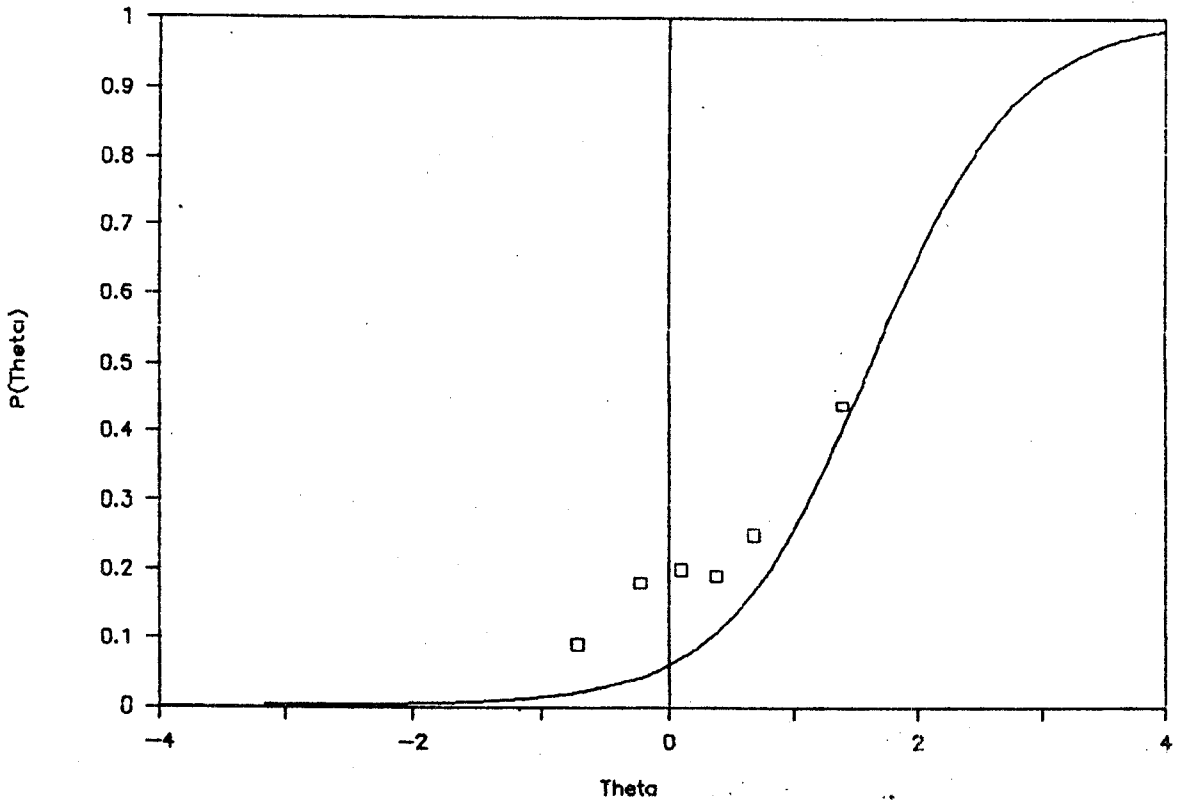
ITEM 04



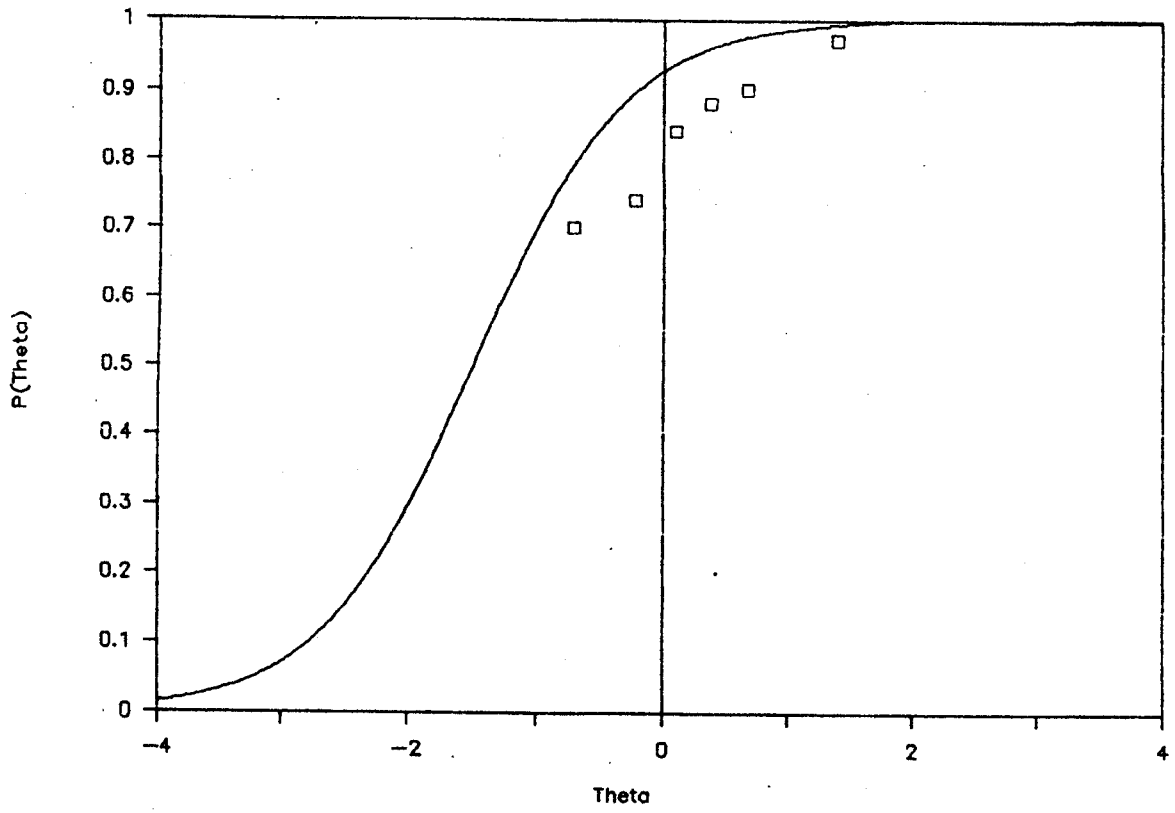
ITEM 05



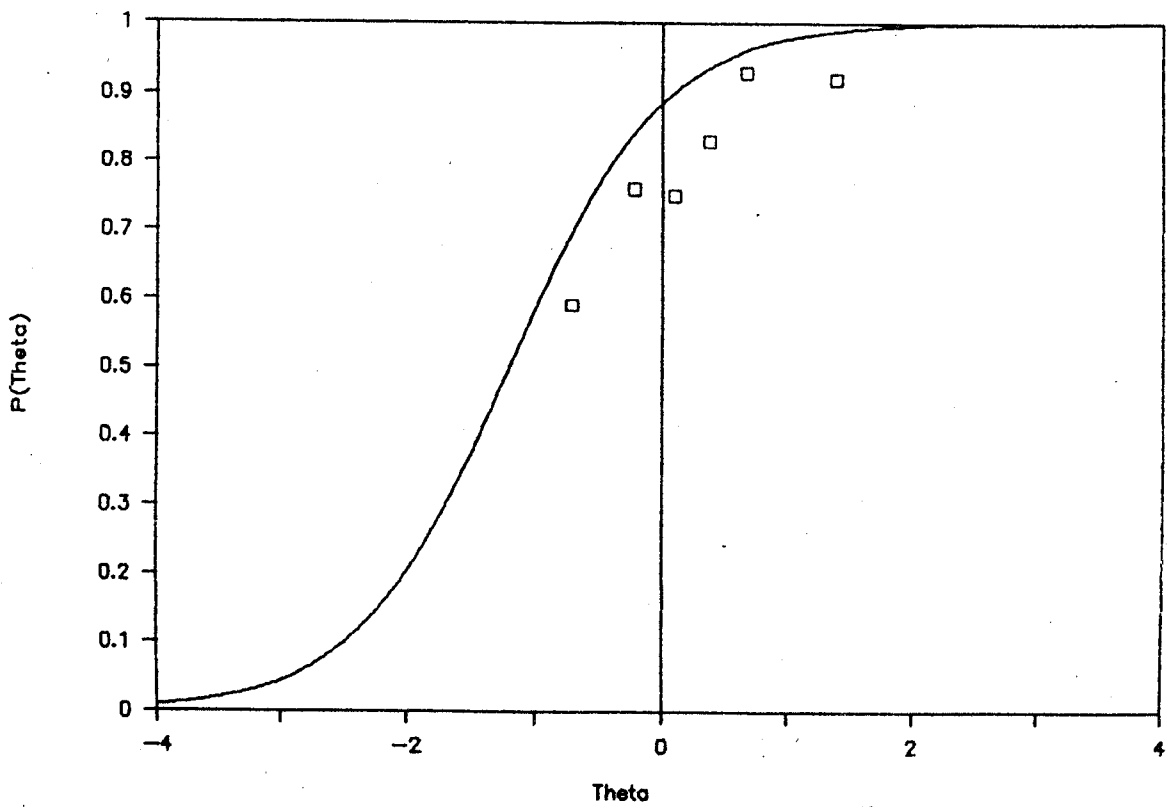
ITEM 06



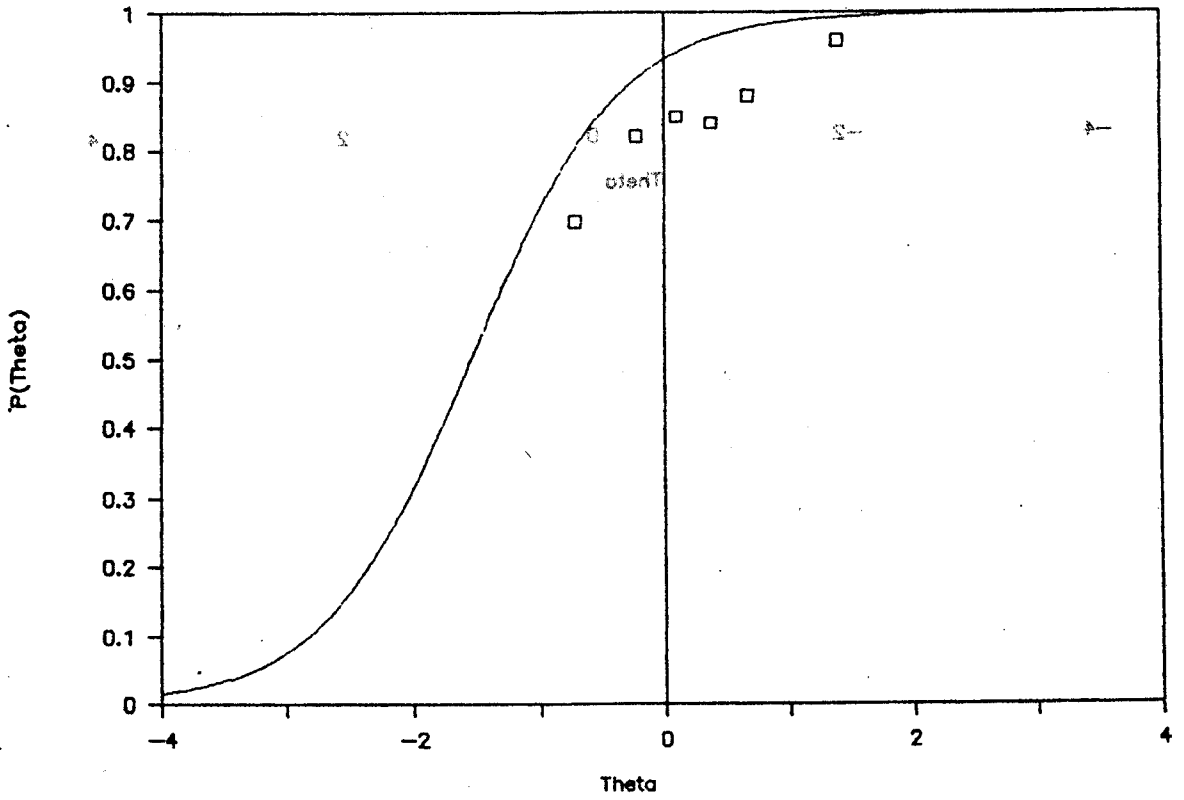
ITEM 07



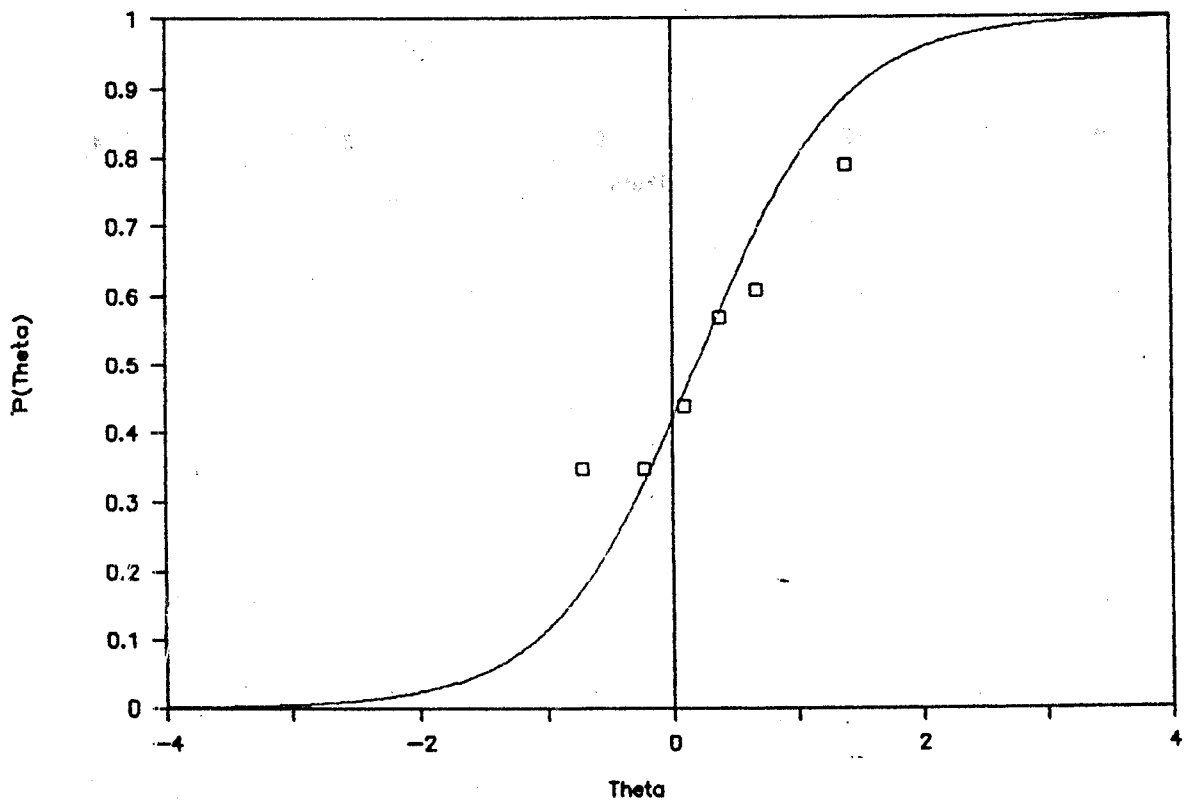
ITEM 08



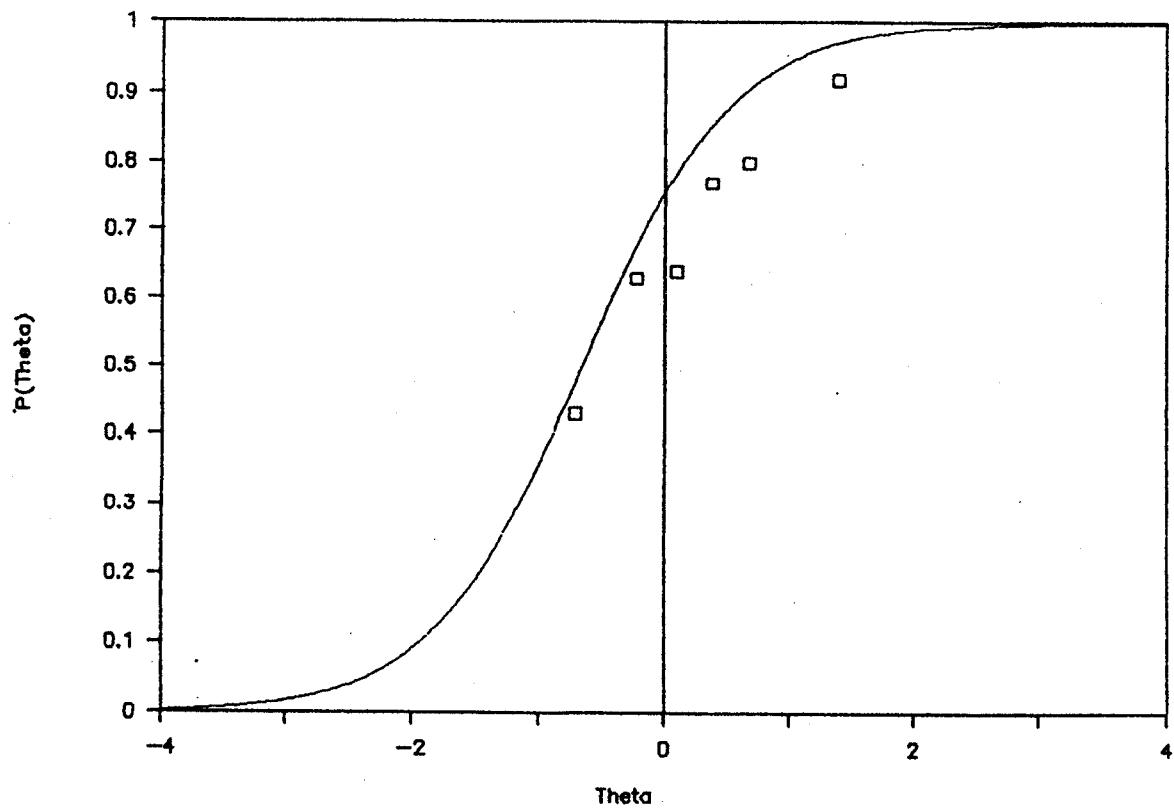
ITEM 09



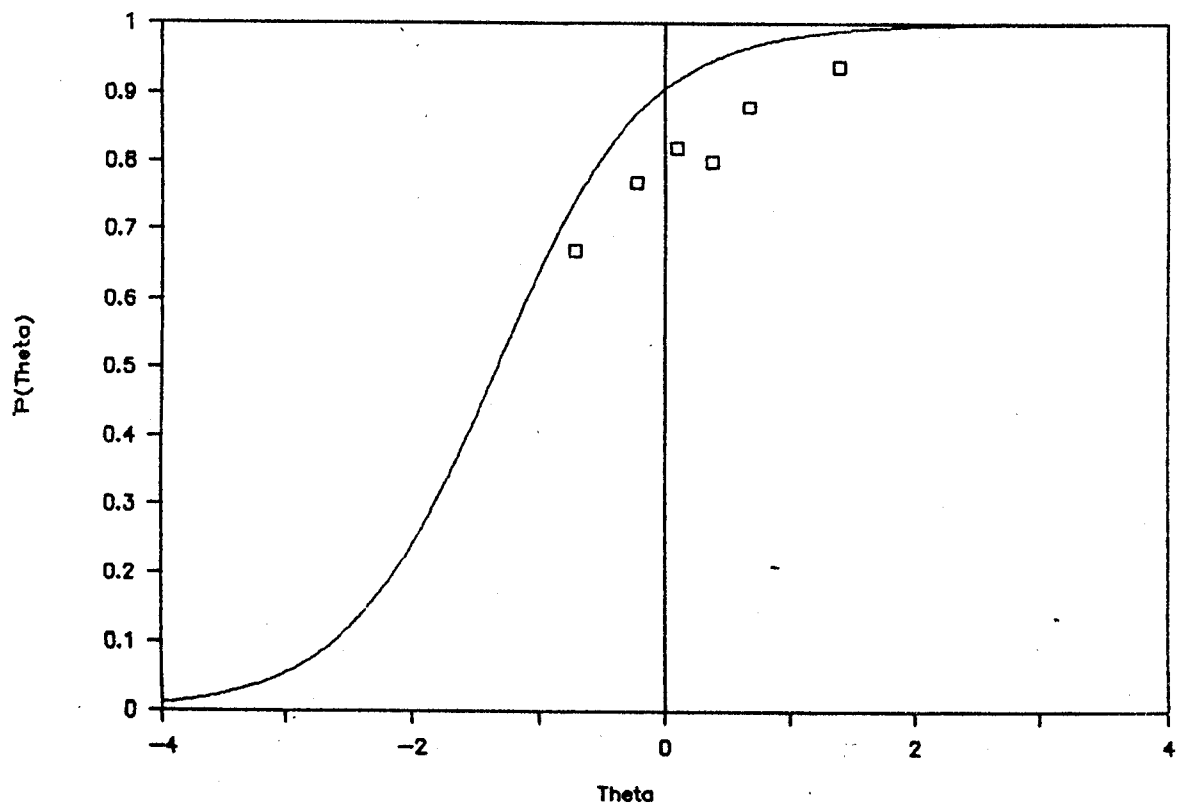
ITEM 10



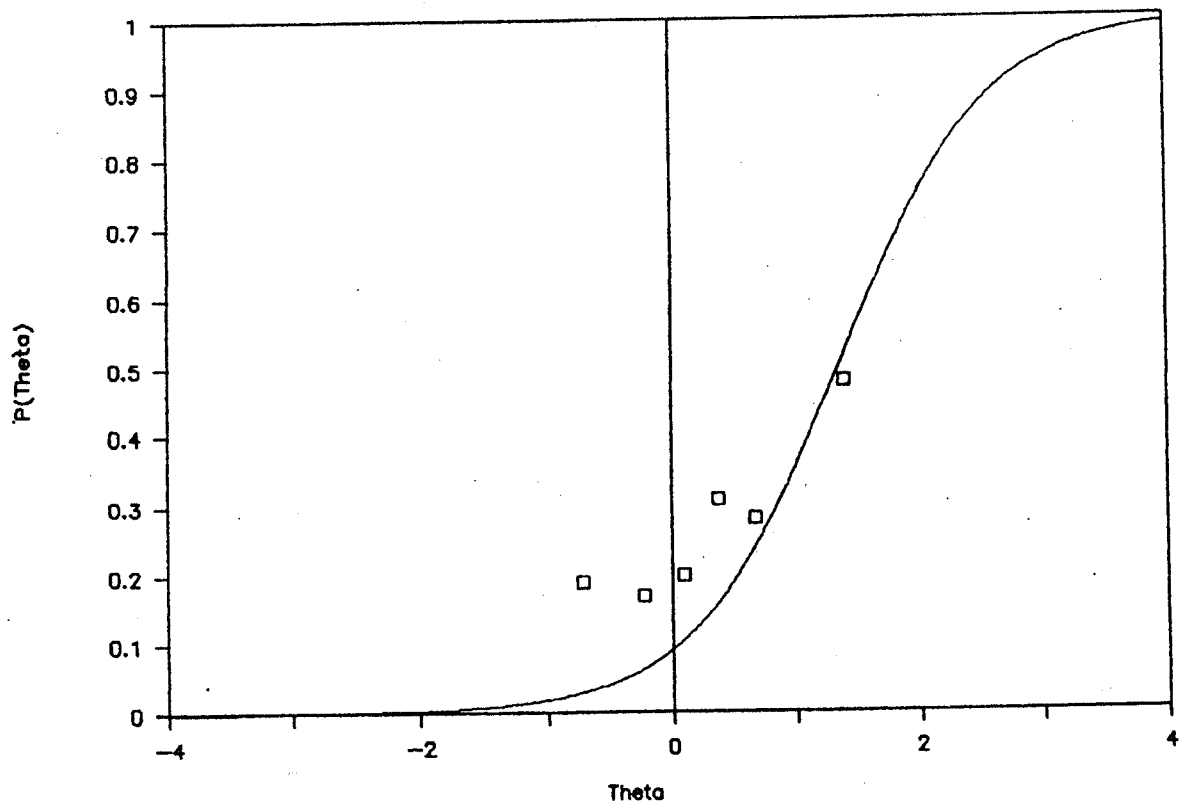
ITEM 11



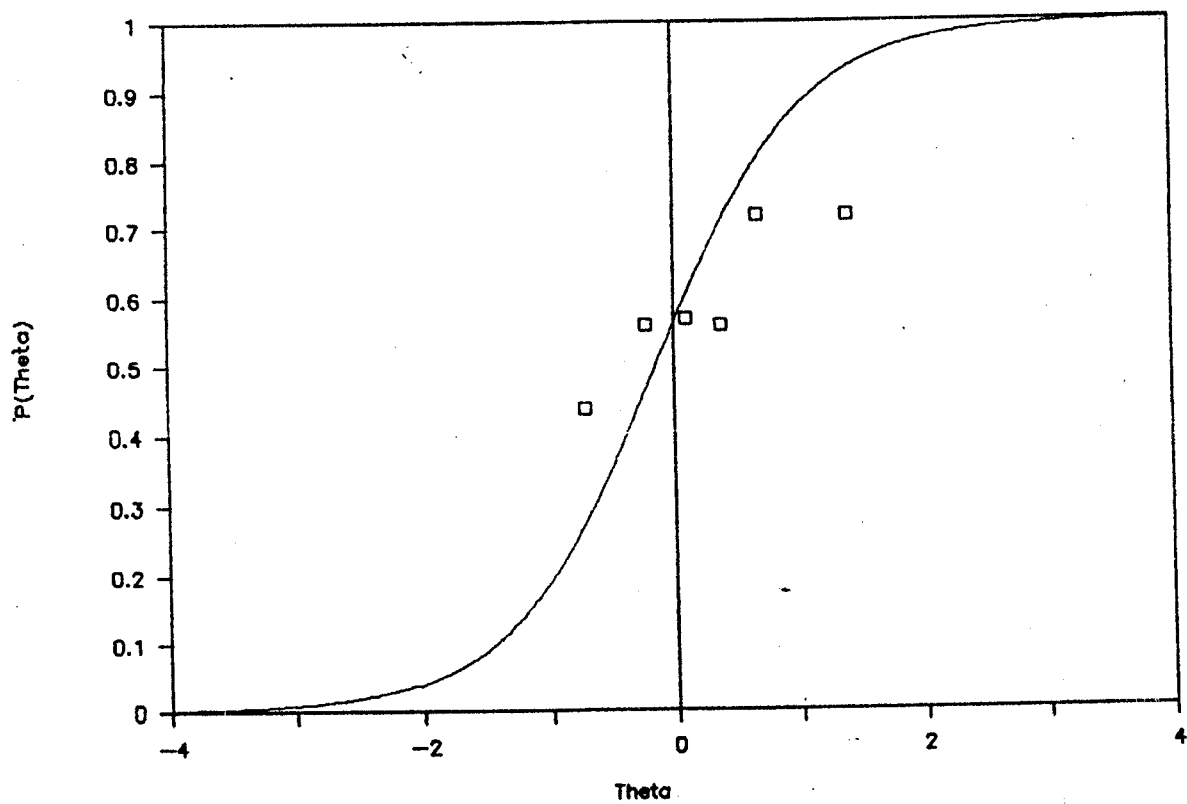
ITEM 12



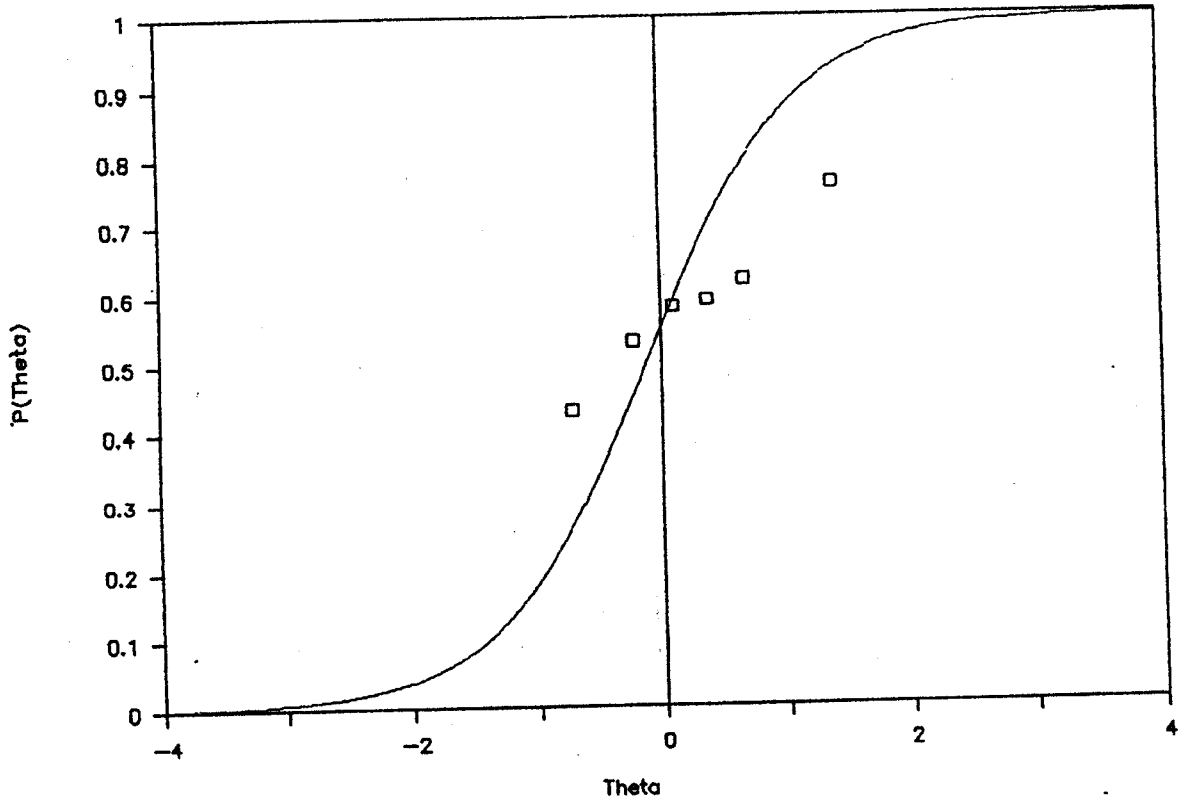
ITEM 13



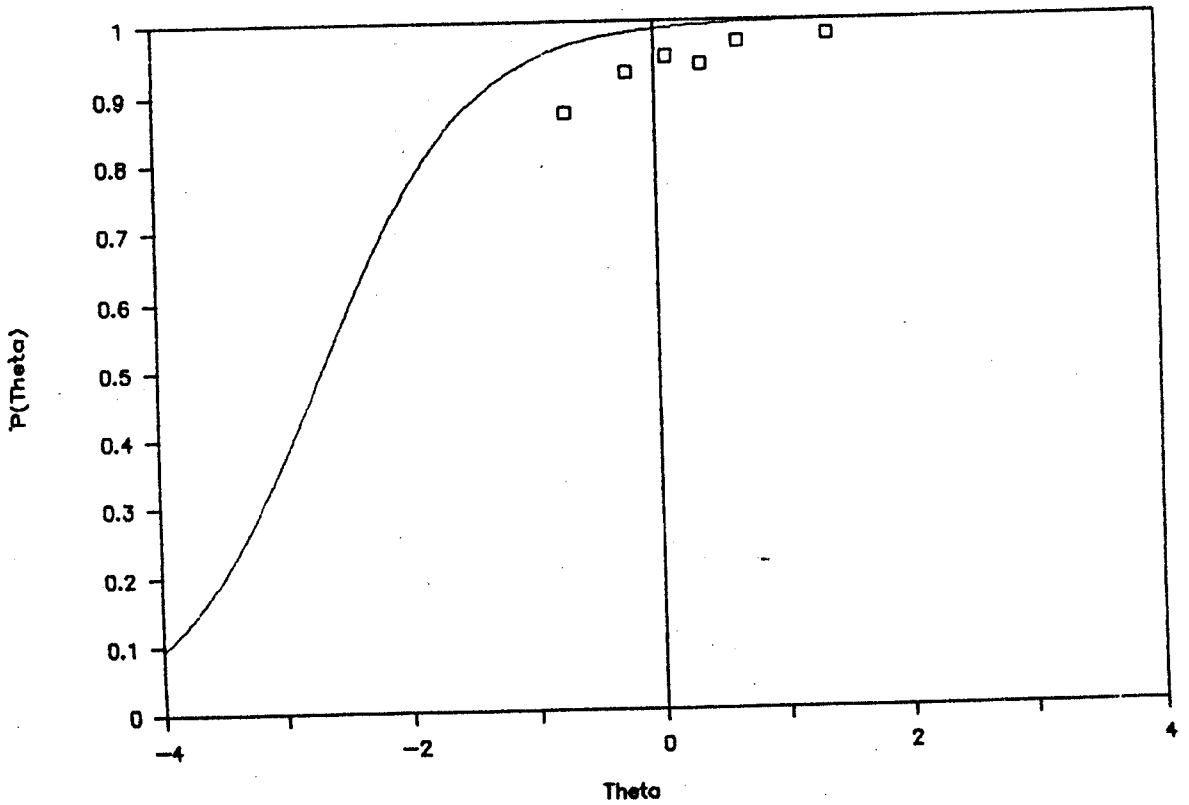
ITEM 15



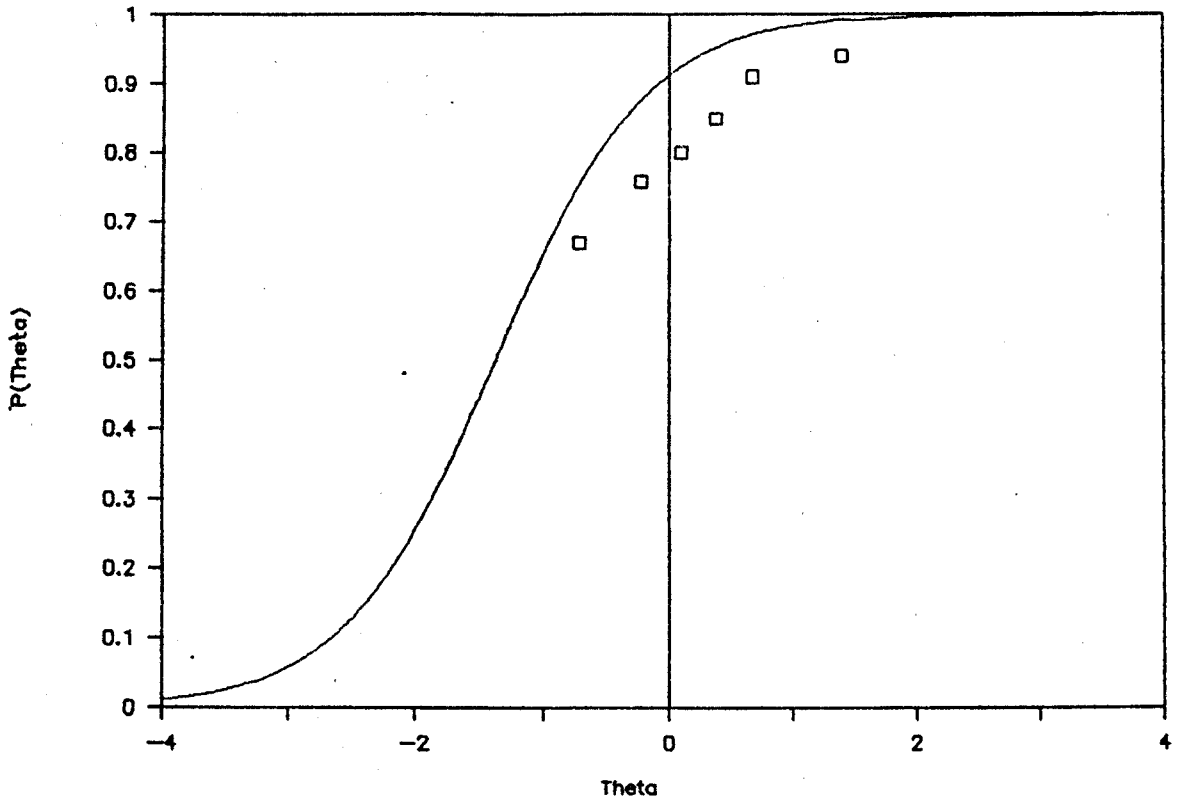
ITEM 16



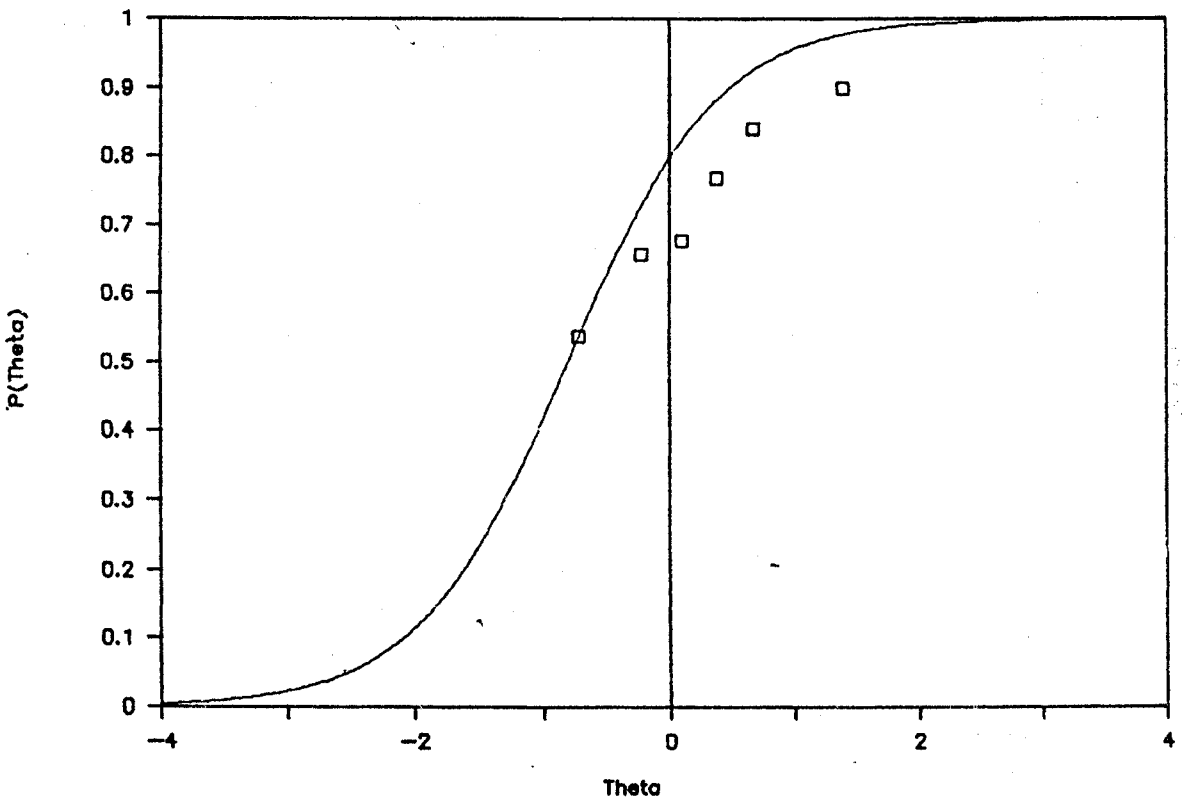
ITEM 17



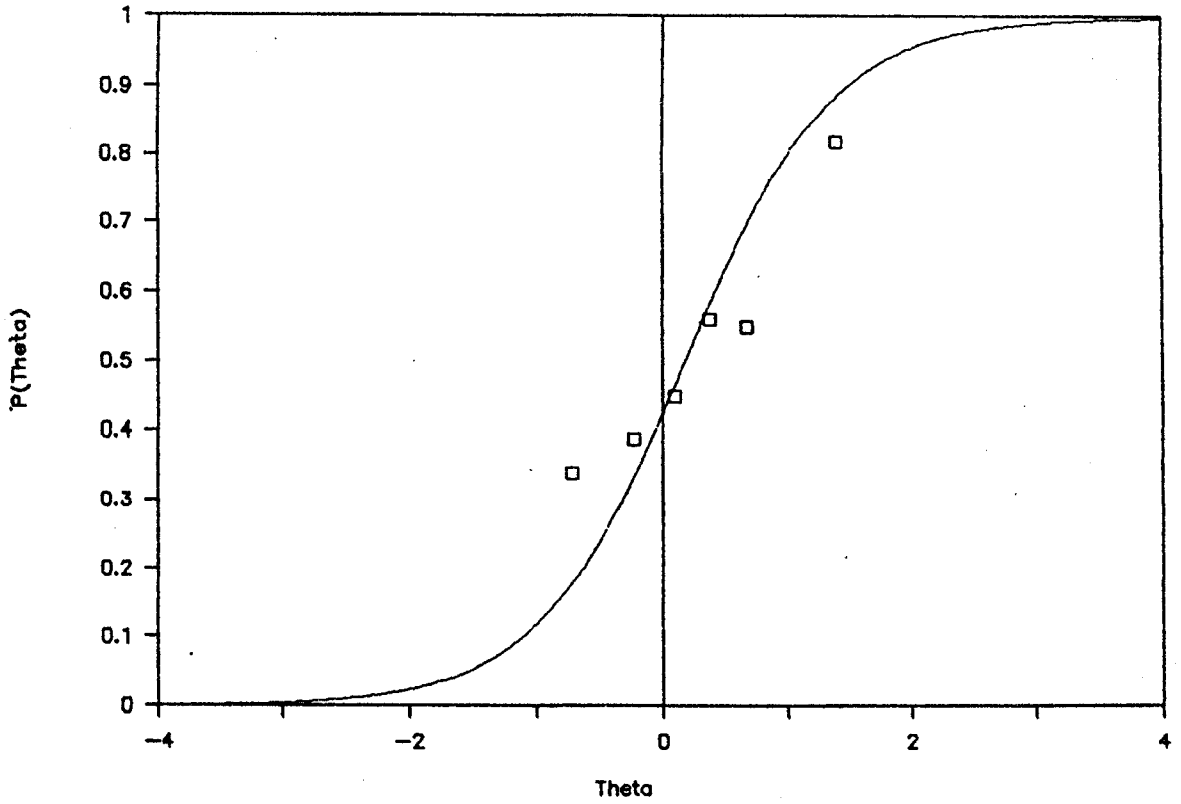
ITEM 18



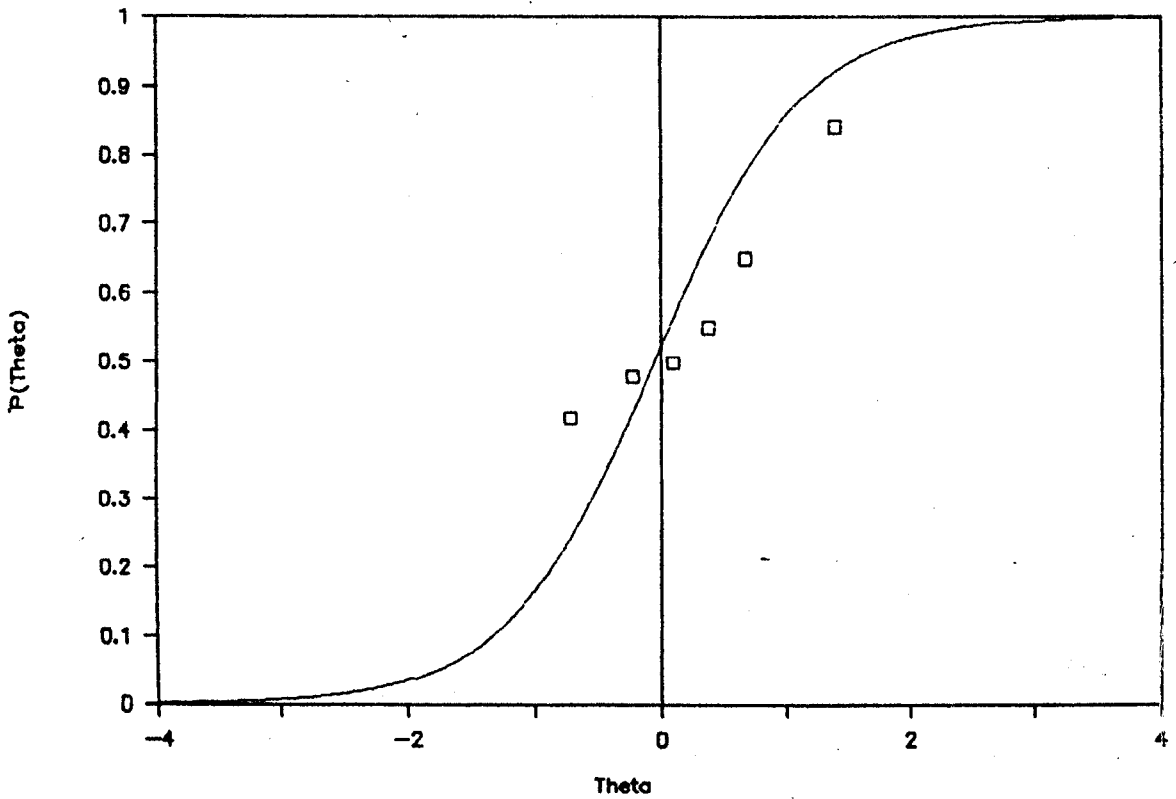
ITEM 19



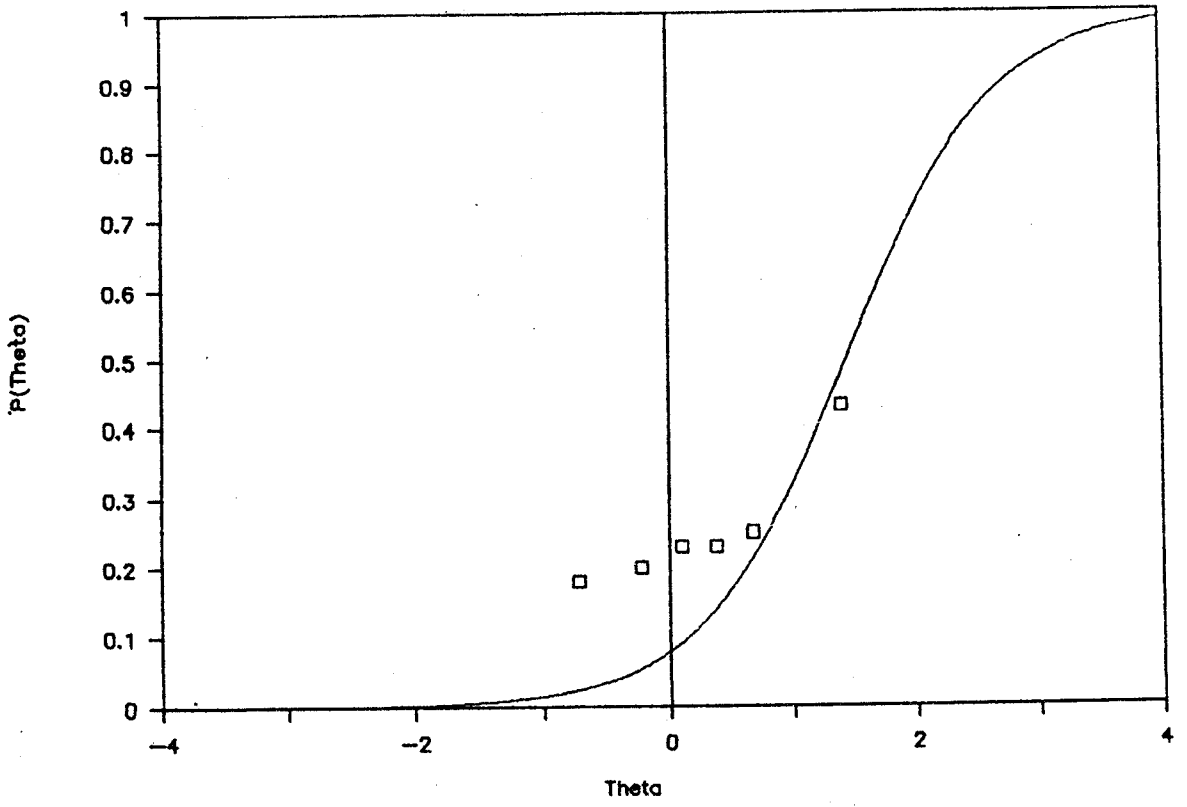
ITEM 20



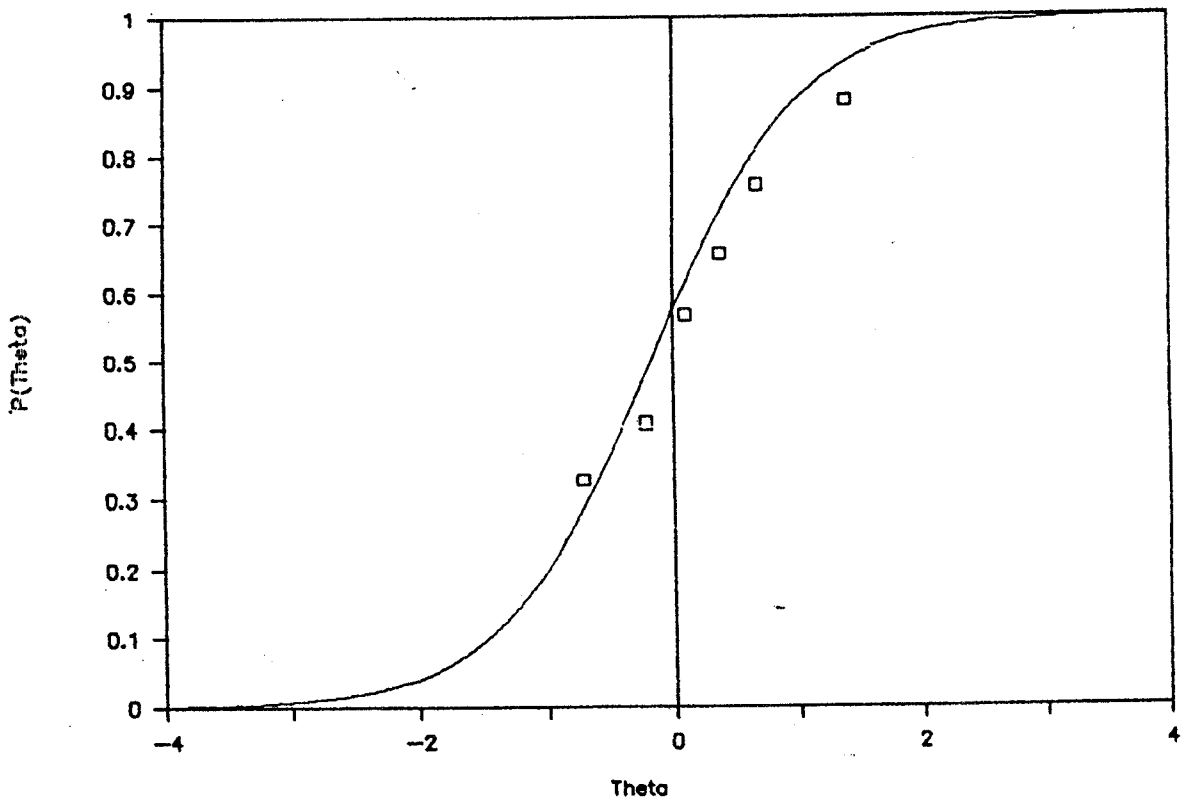
ITEM 21



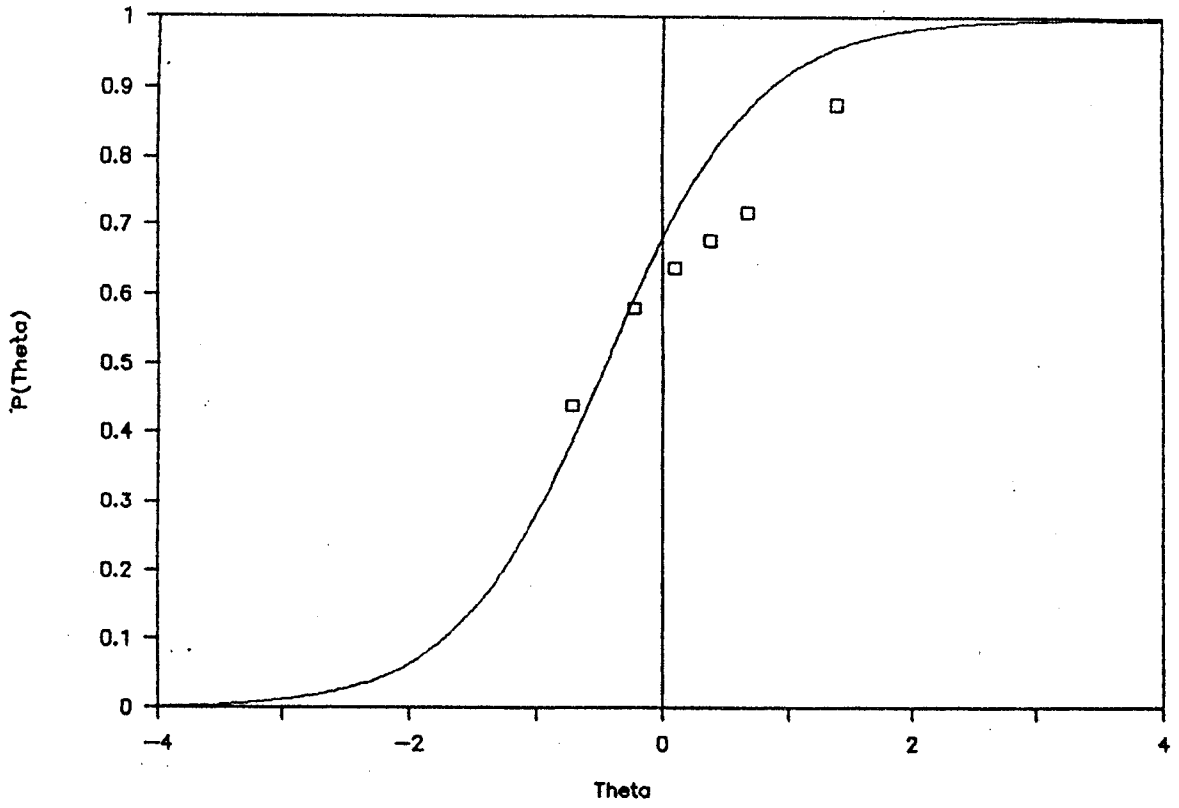
ITEM 22



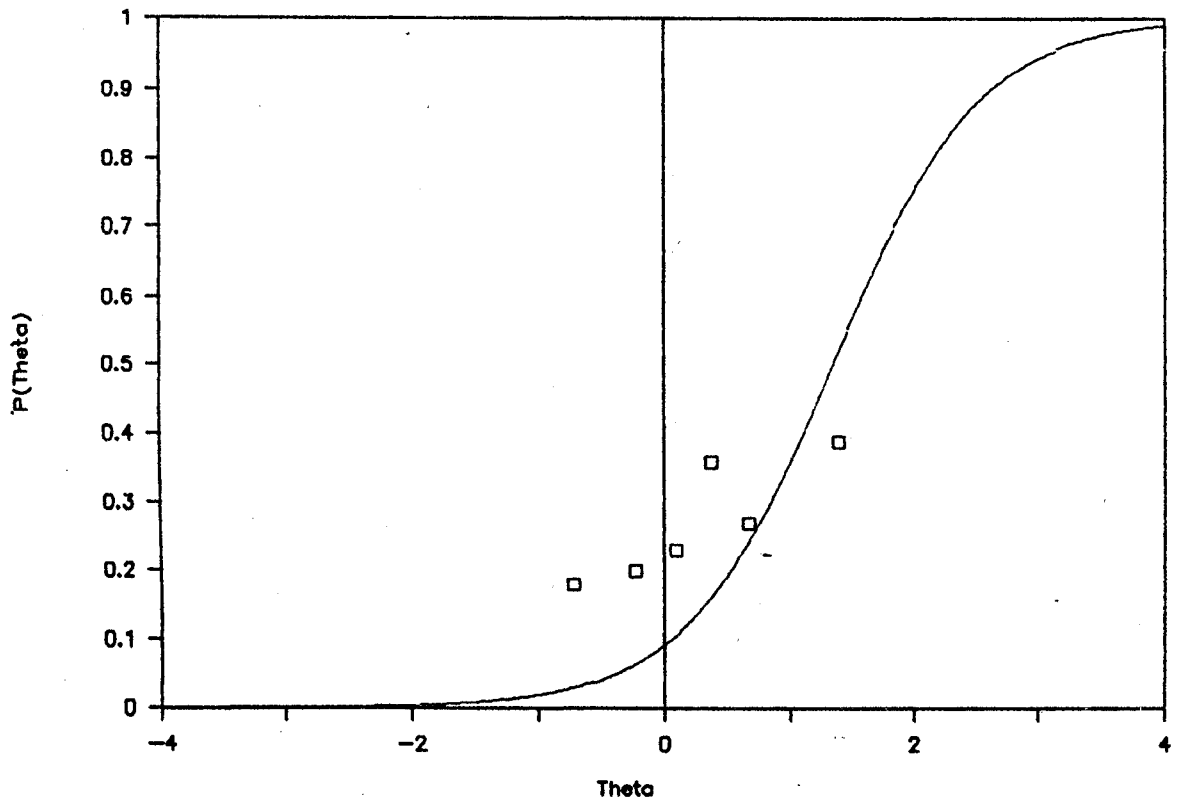
ITEM 23



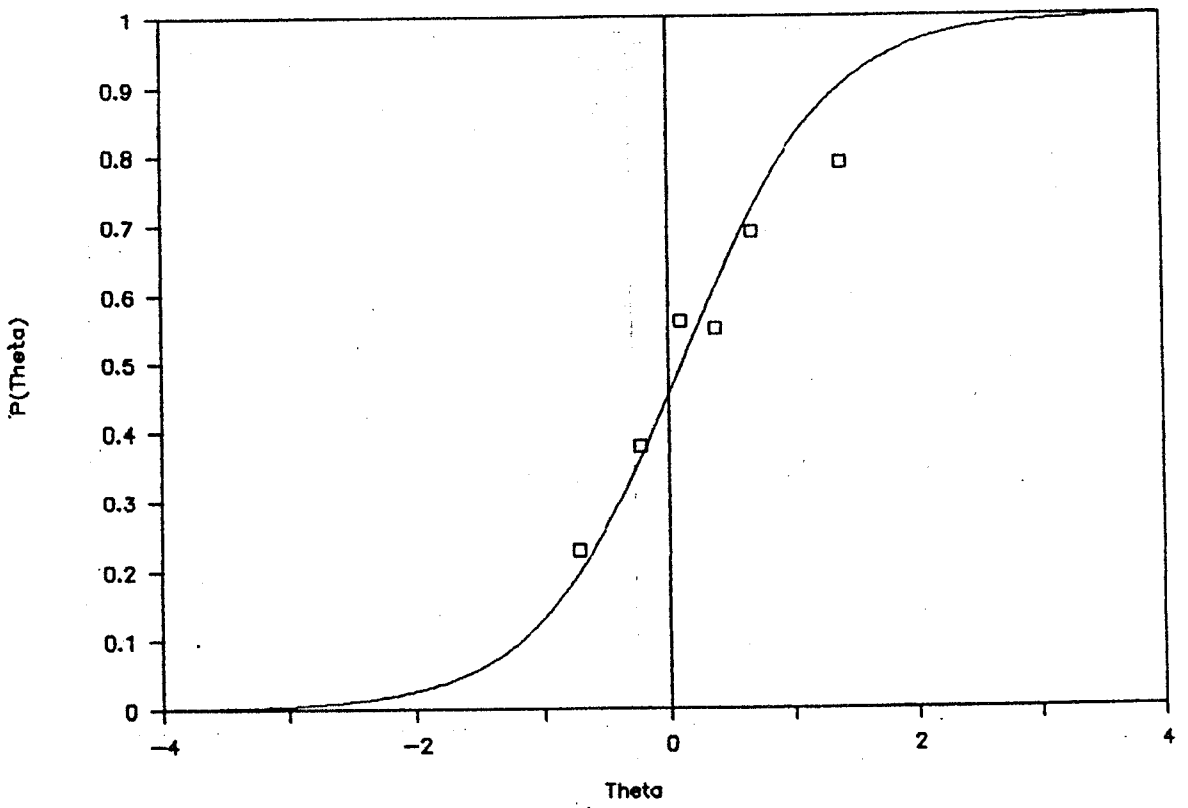
ITEM 24



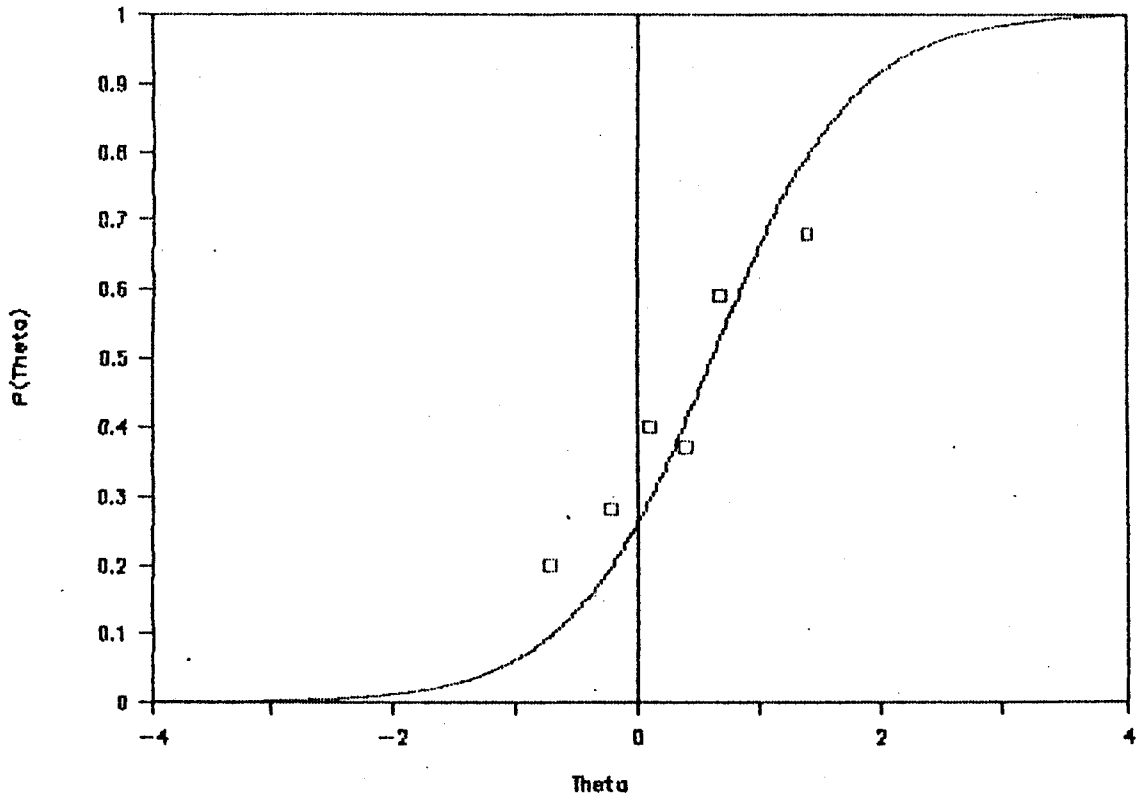
ITEM 25



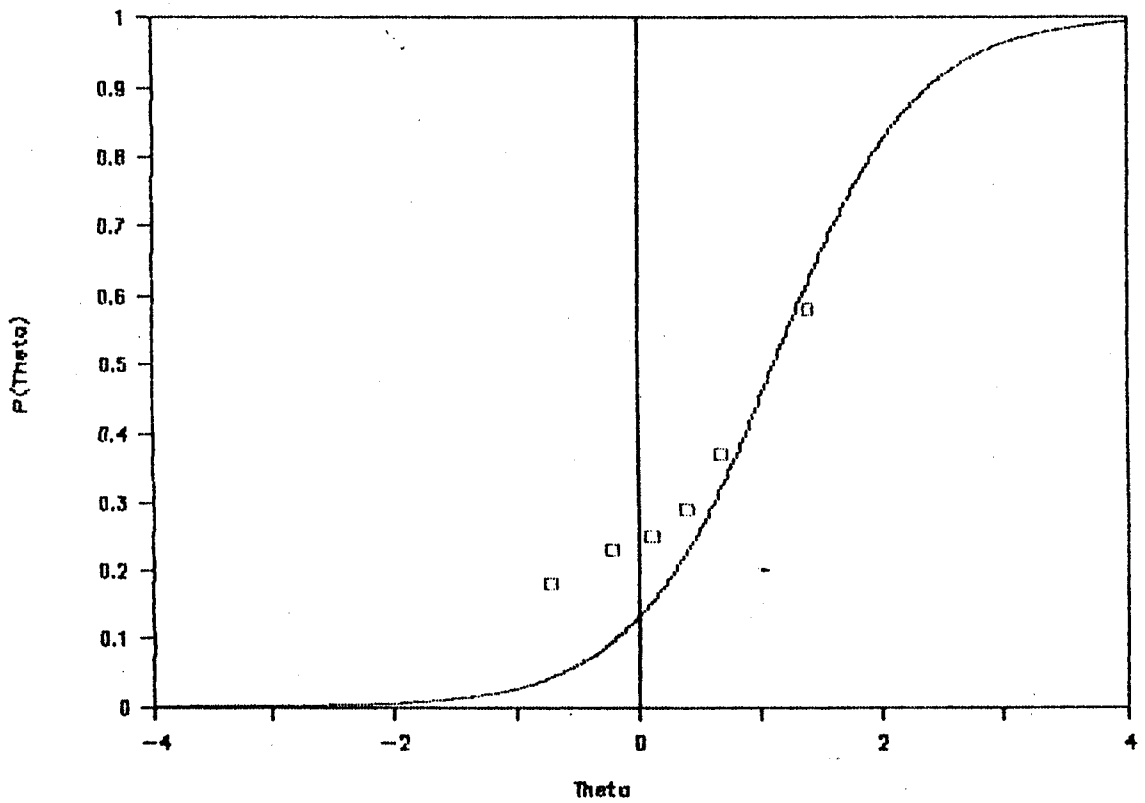
ITEM 26



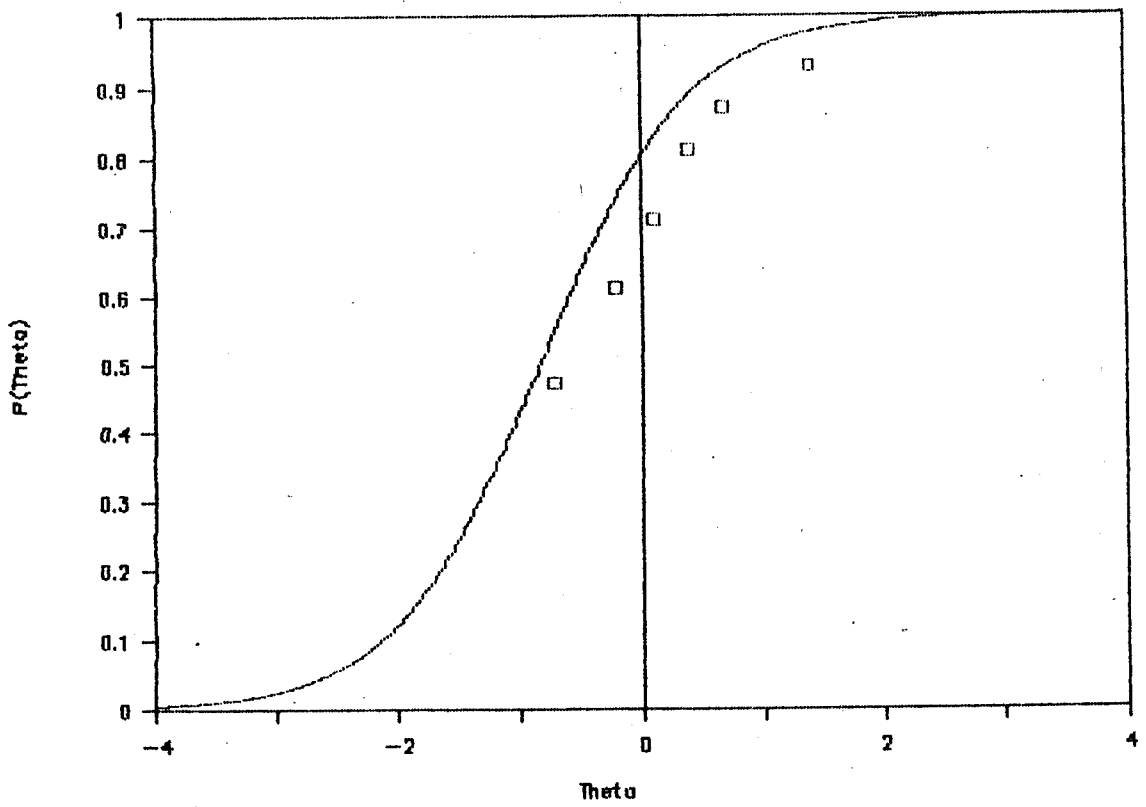
ITEM 27



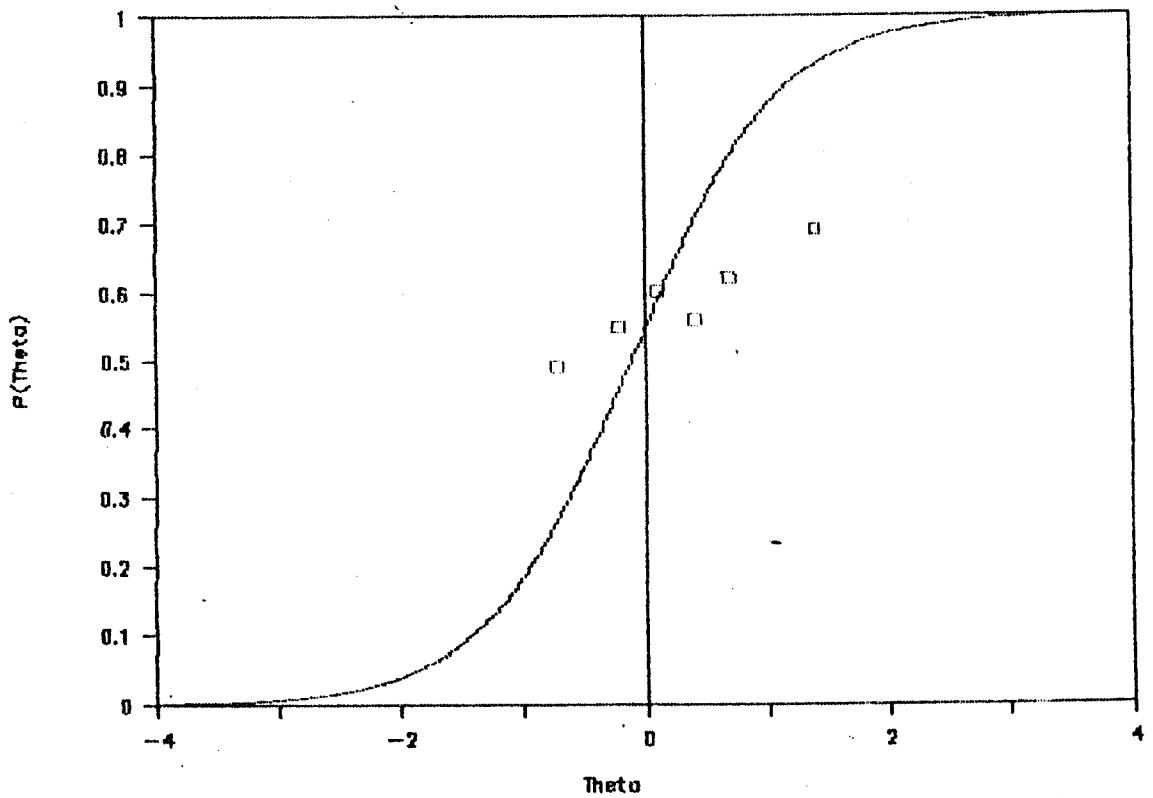
ITEM 28



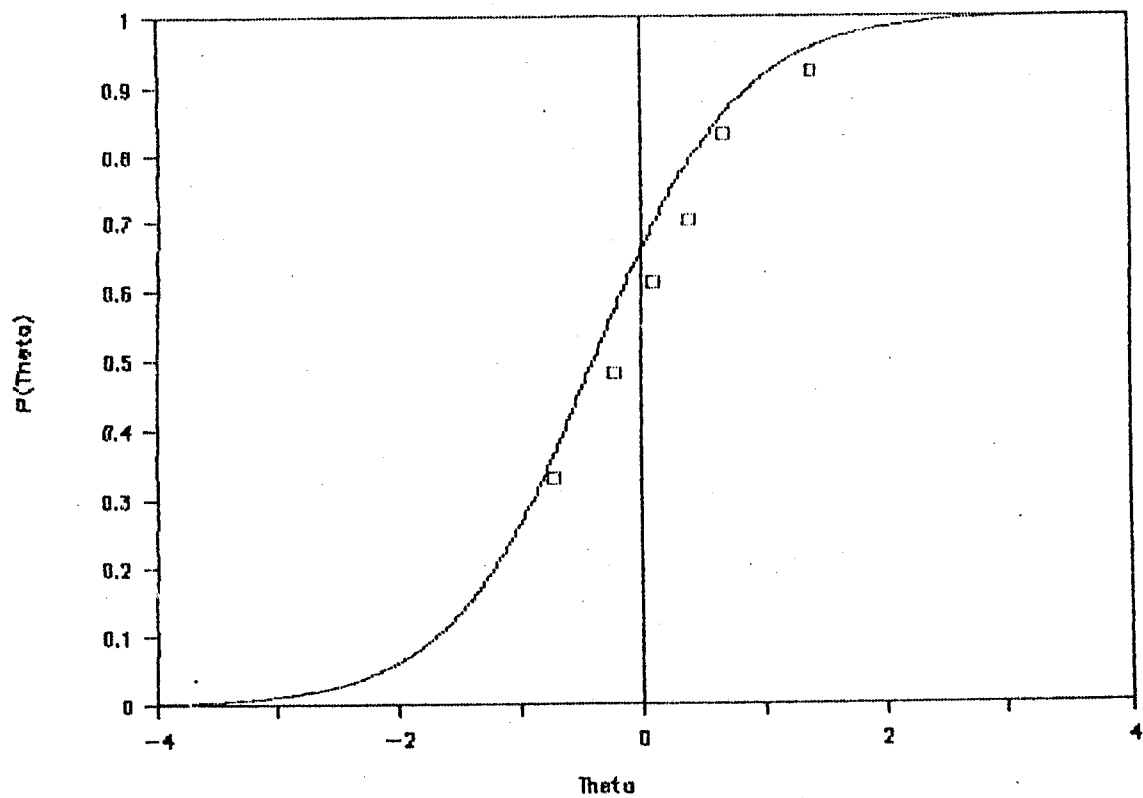
ITEM 29



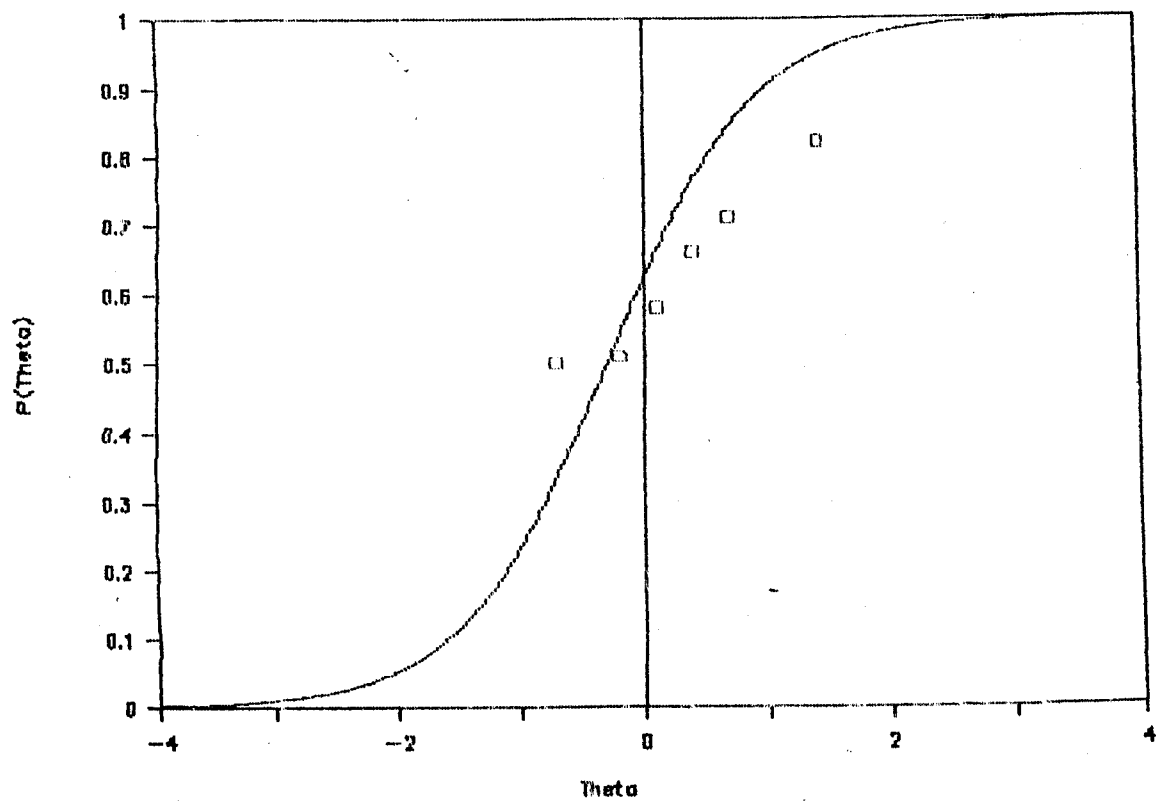
ITEM 30



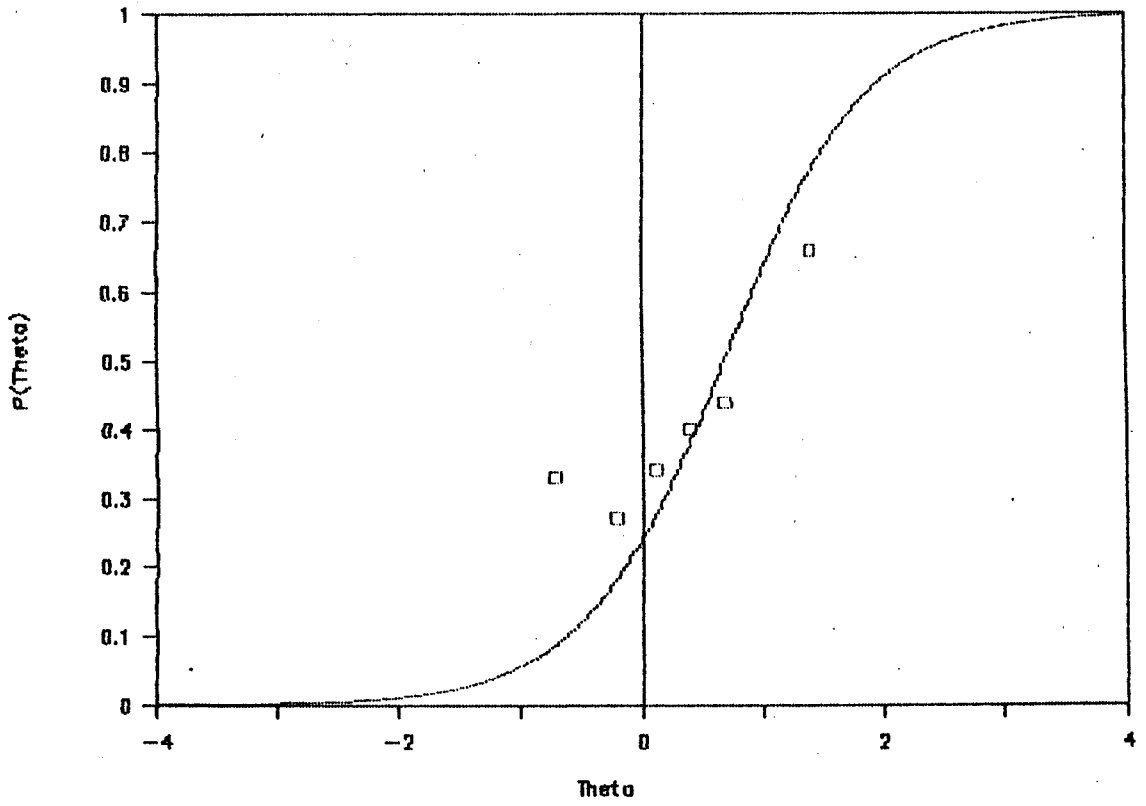
ITEM 31



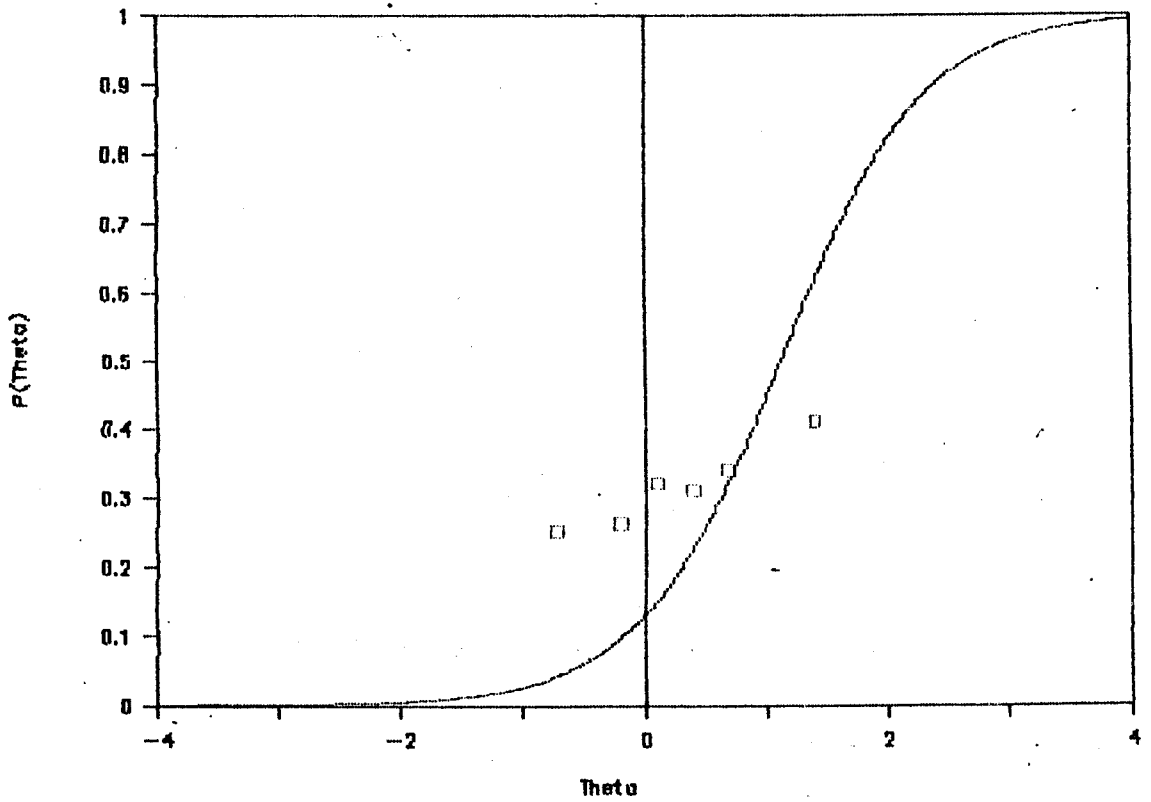
ITEM 32



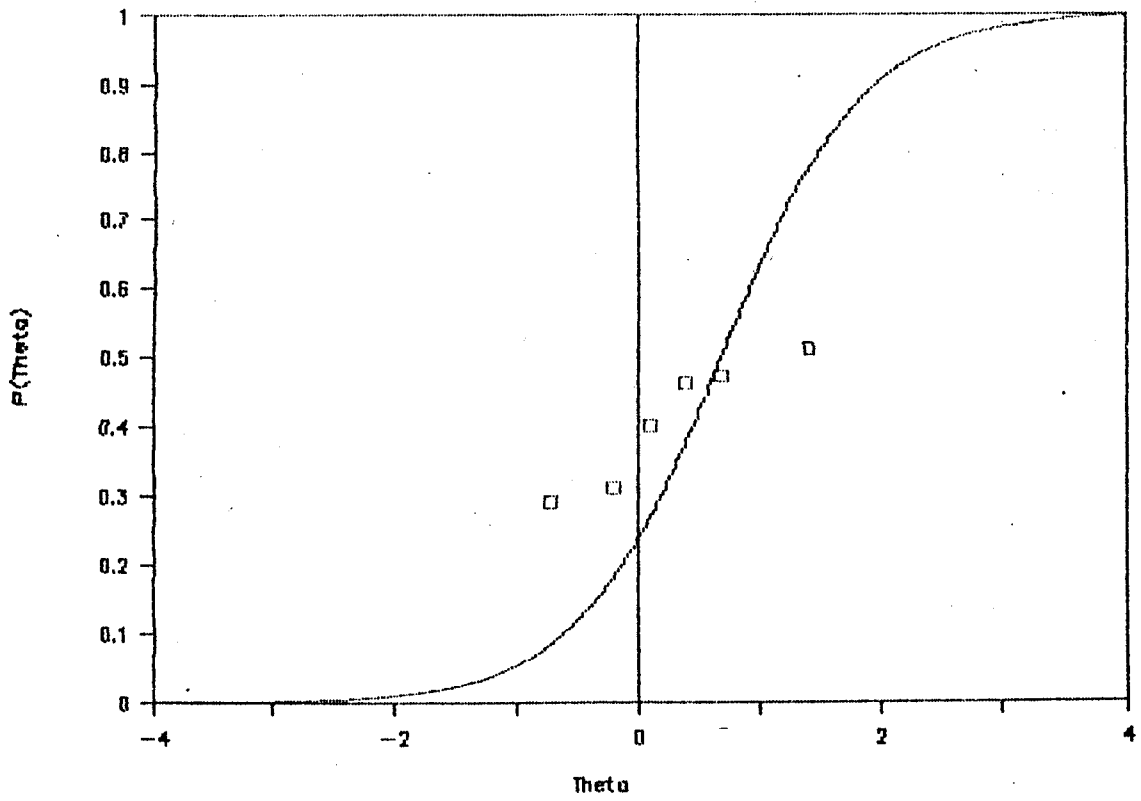
ITEM 33



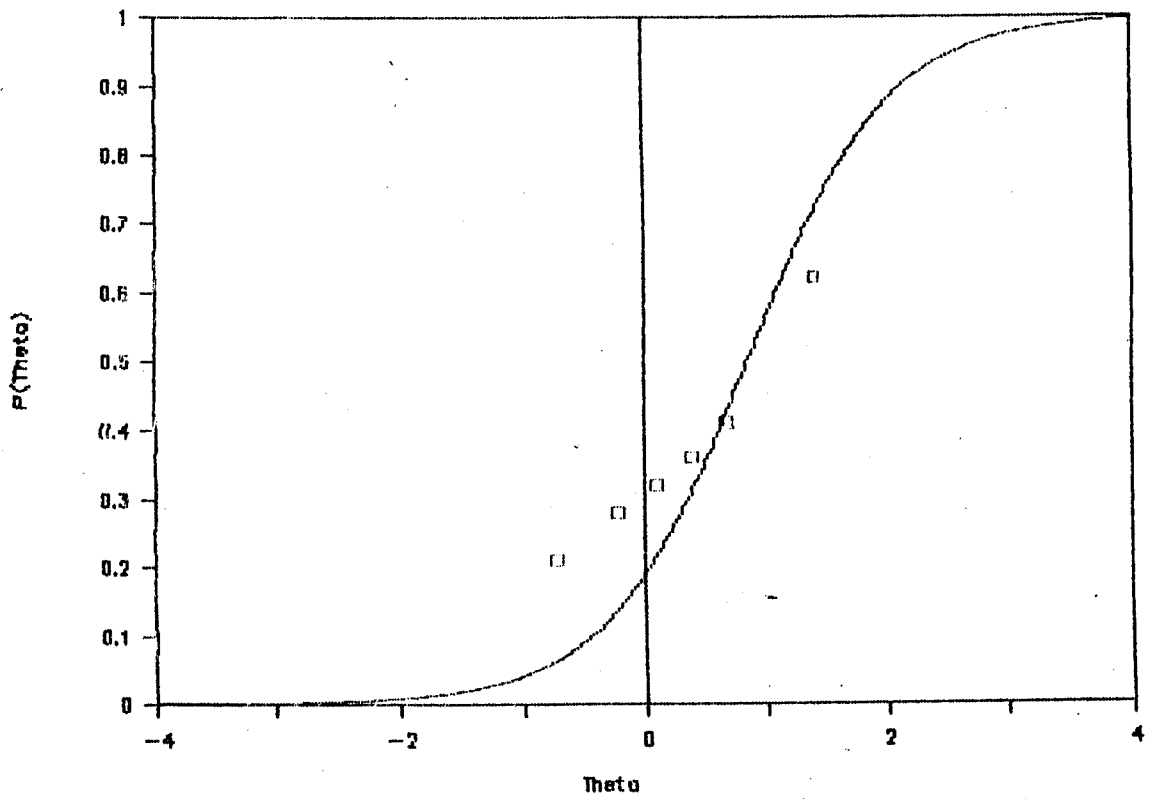
ITEM 34



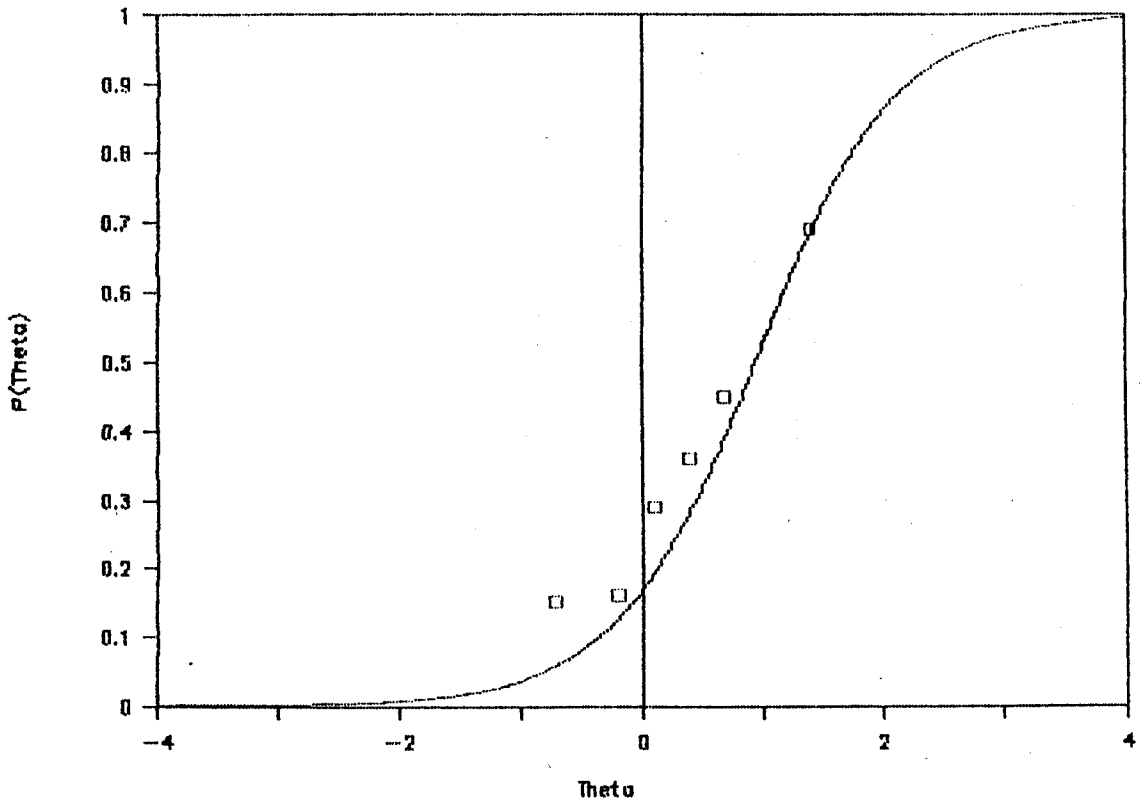
ITEM 35



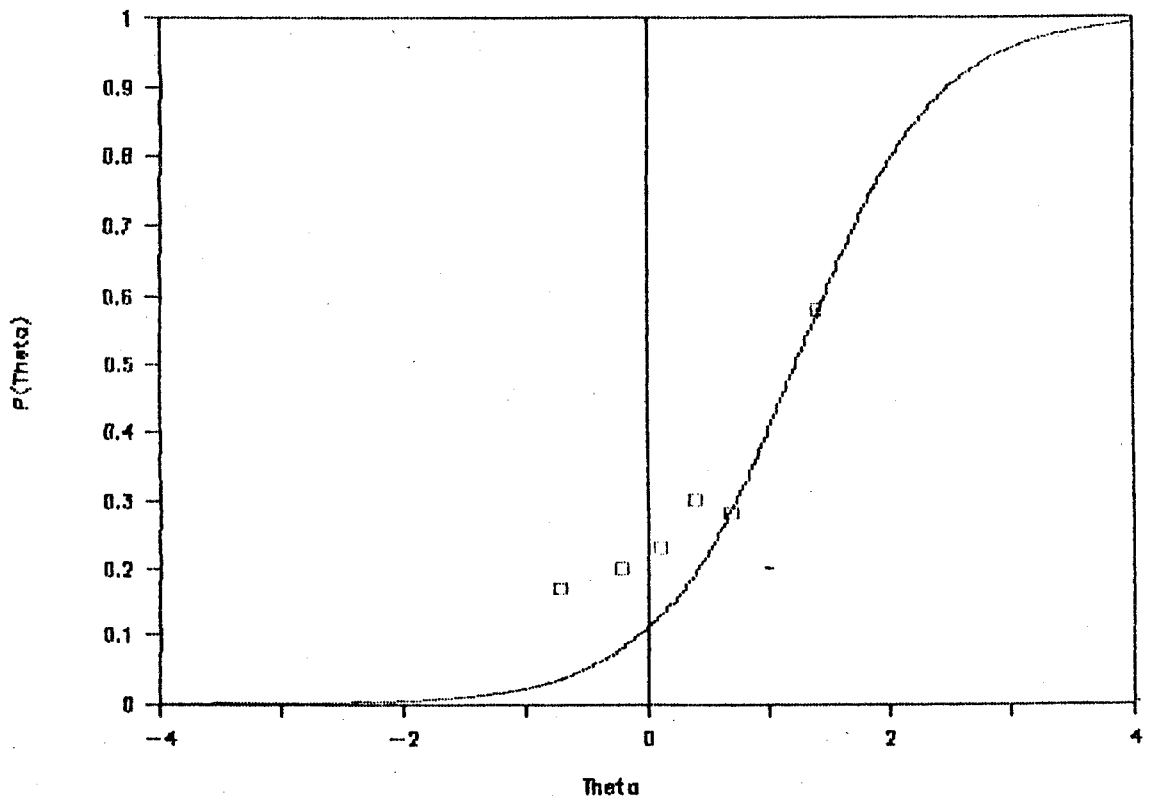
ITEM 36



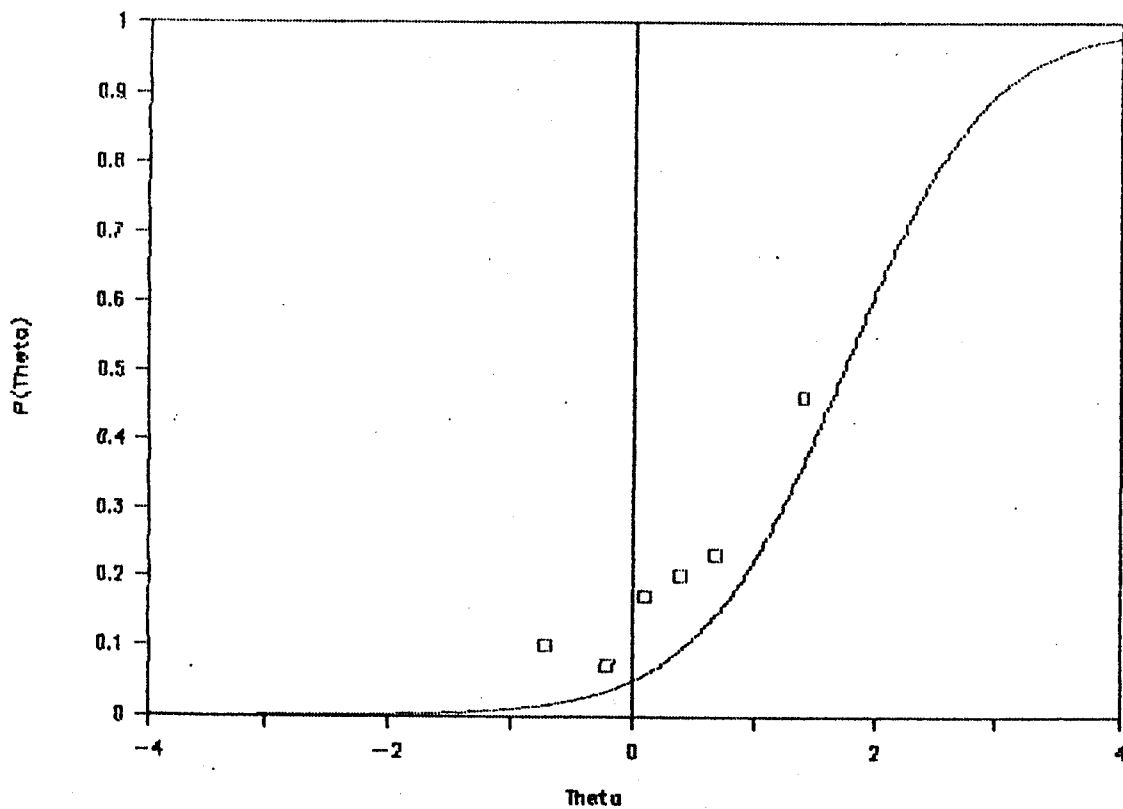
ITEM 37



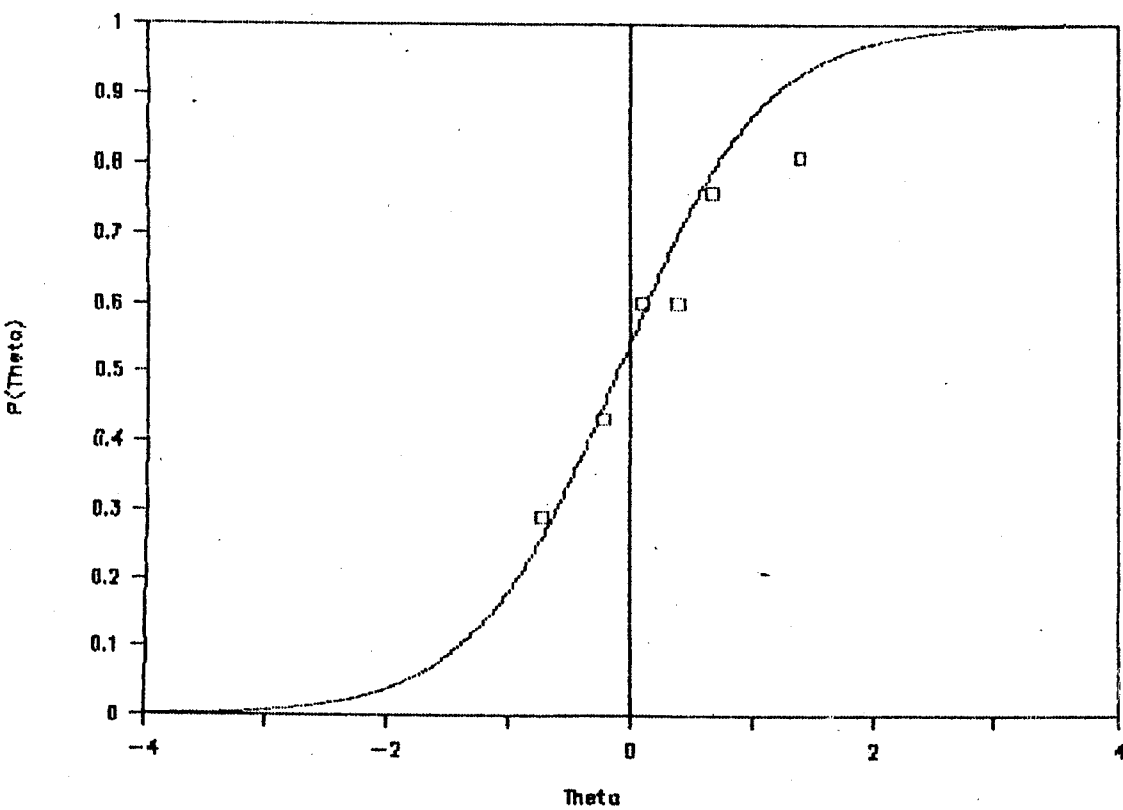
ITEM 38



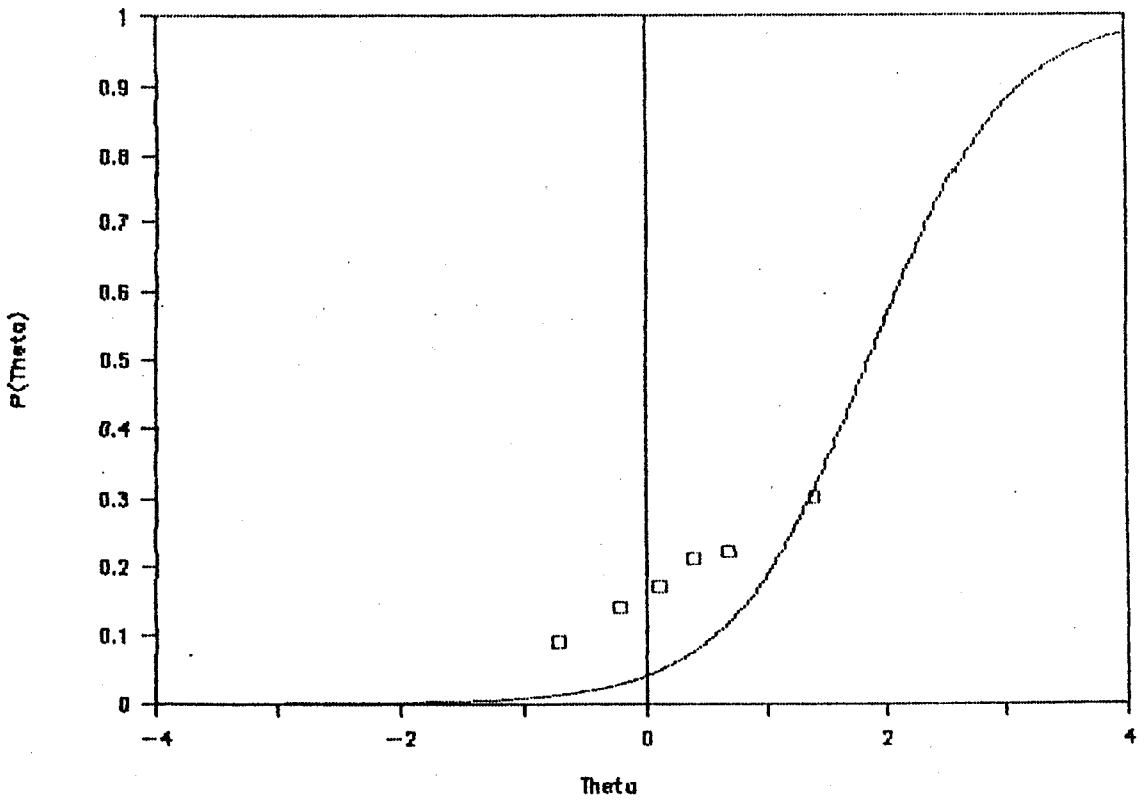
ITEM 39



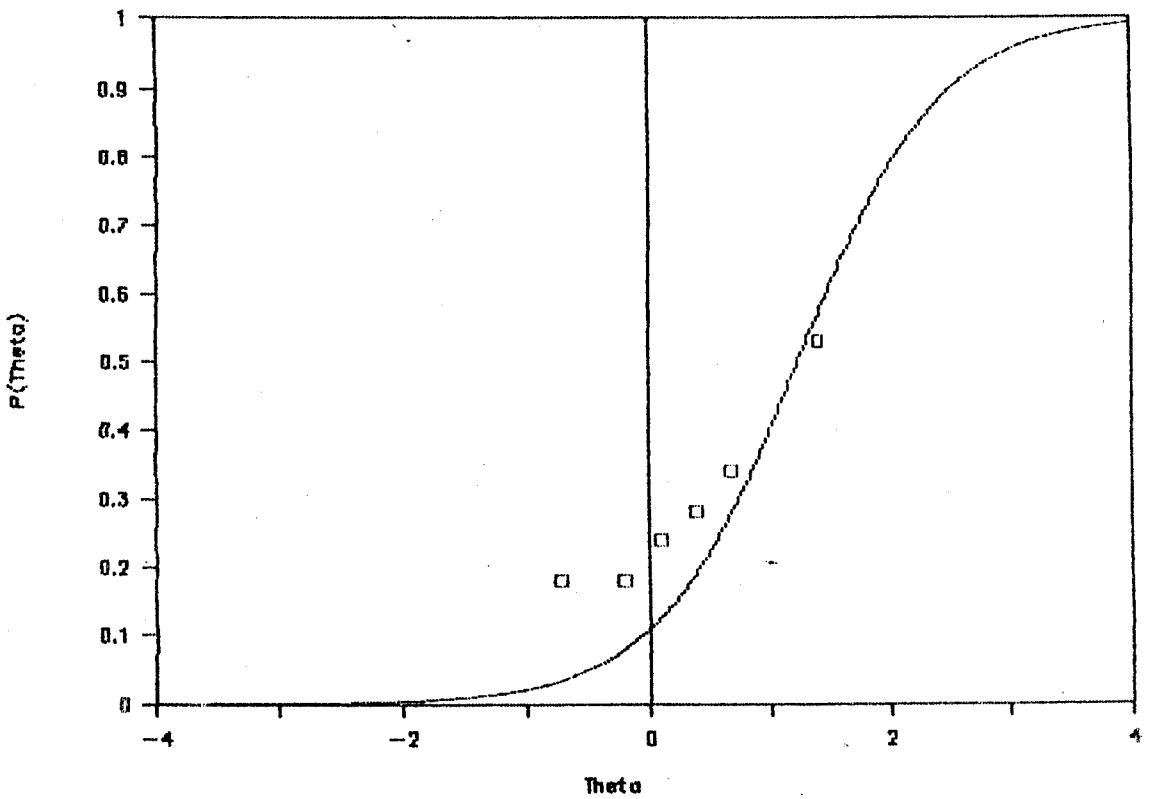
ITEM 40



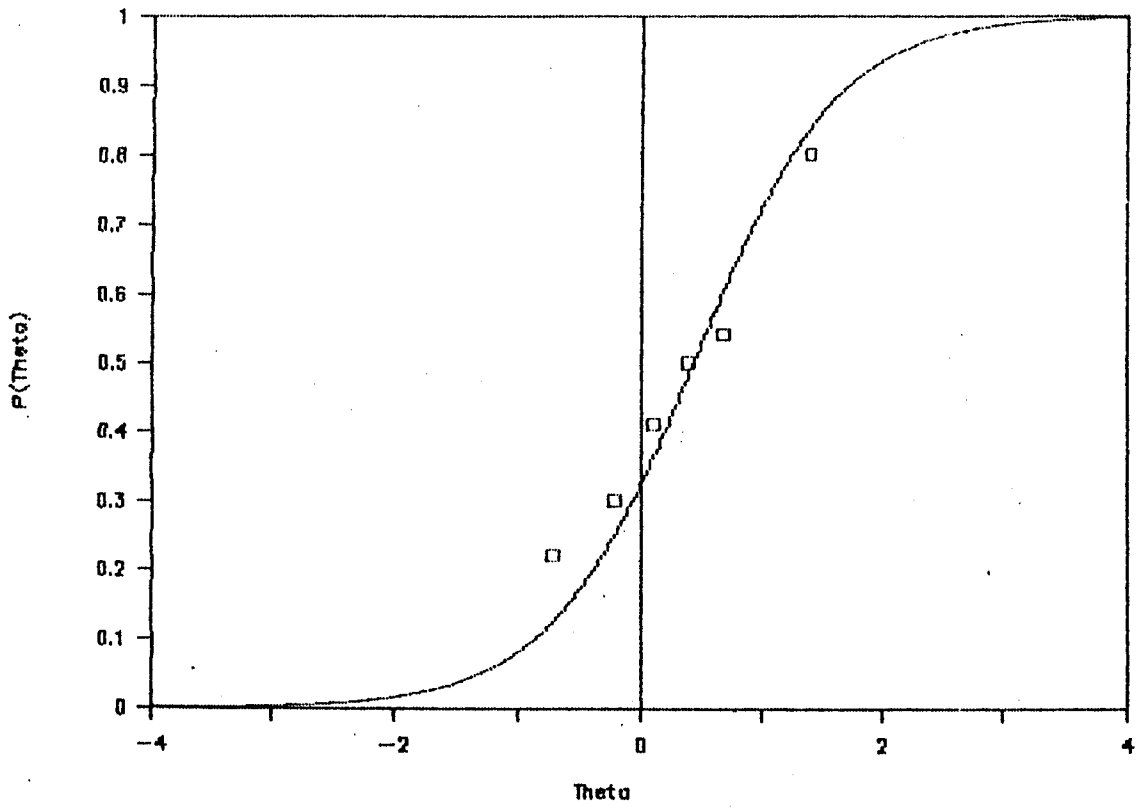
ITEM 41



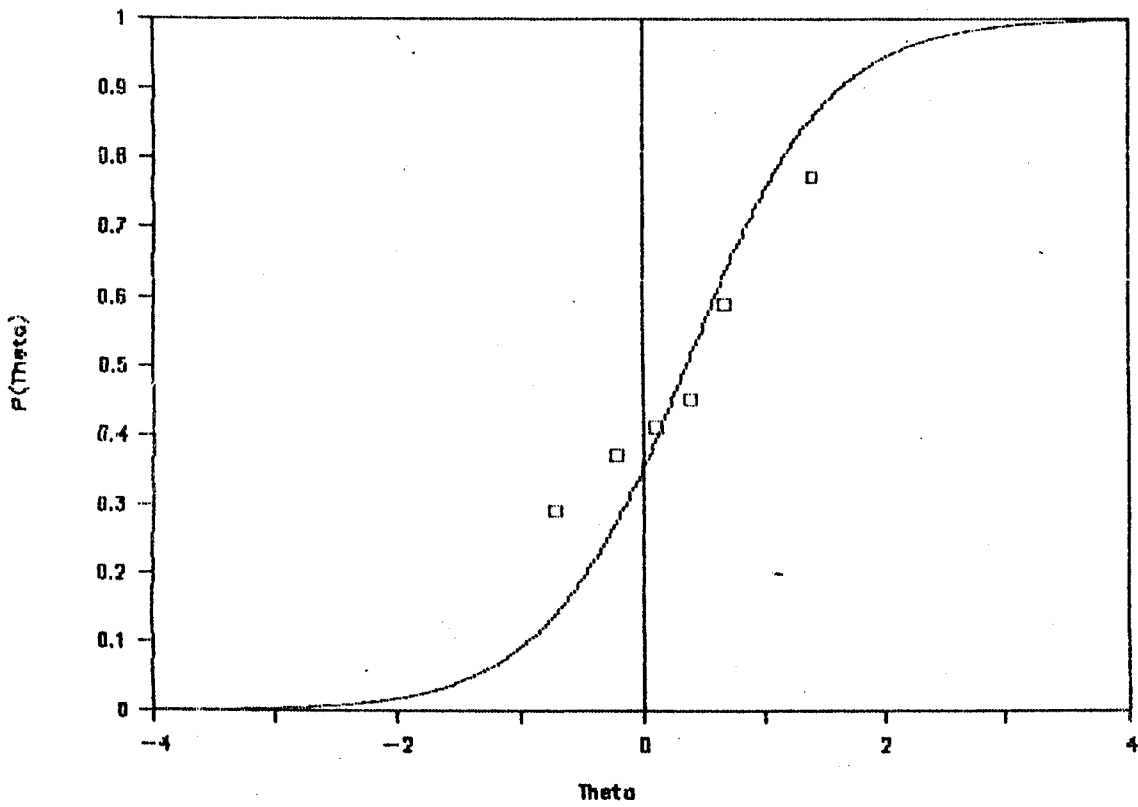
ITEM 42



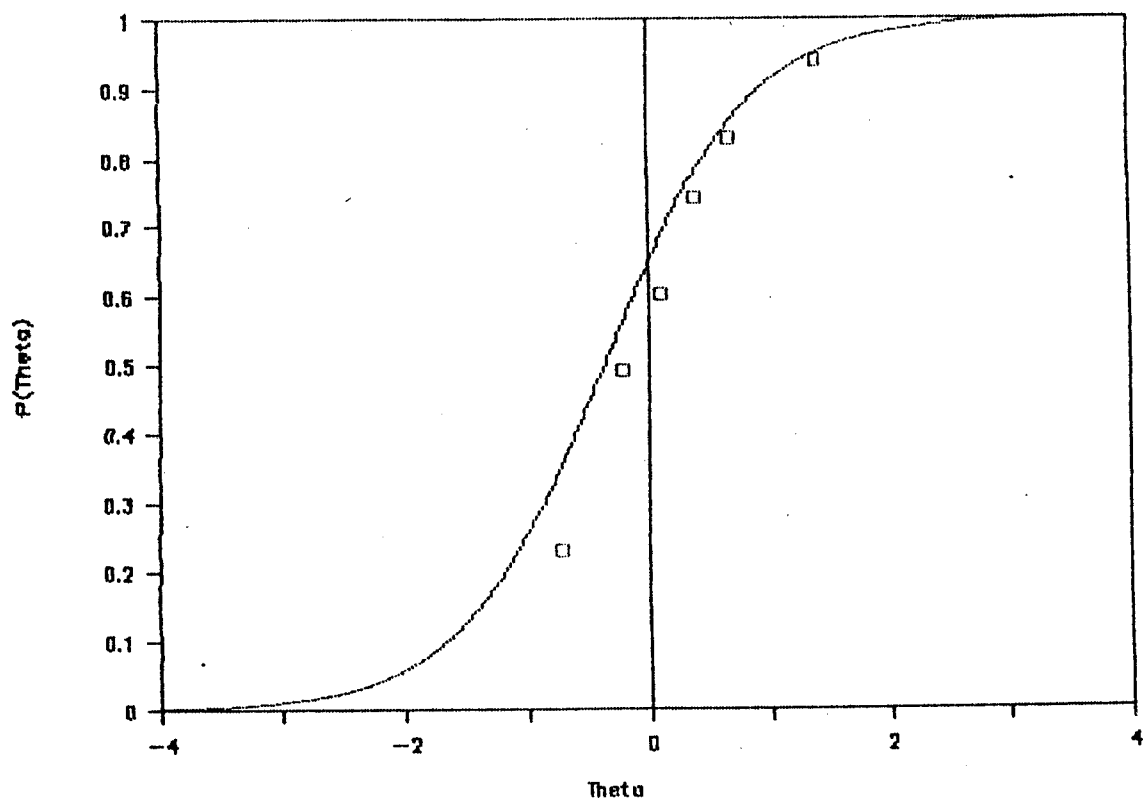
ITEM 43



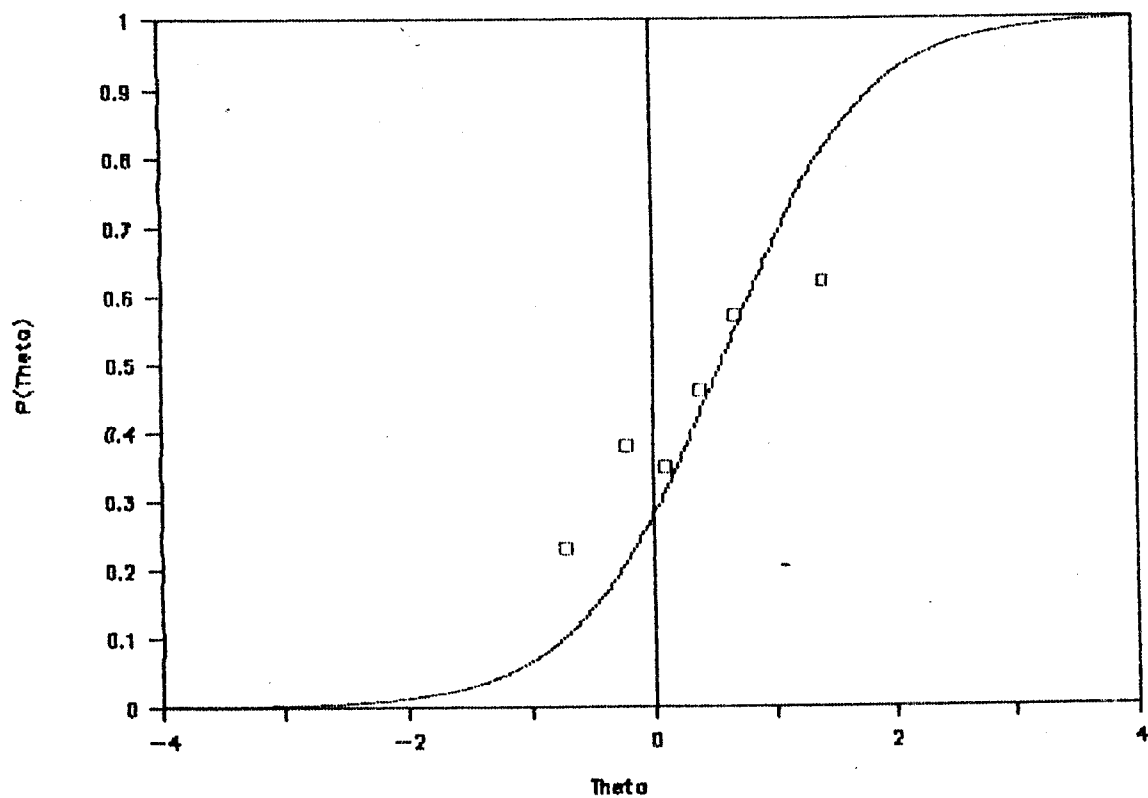
ITEM 44



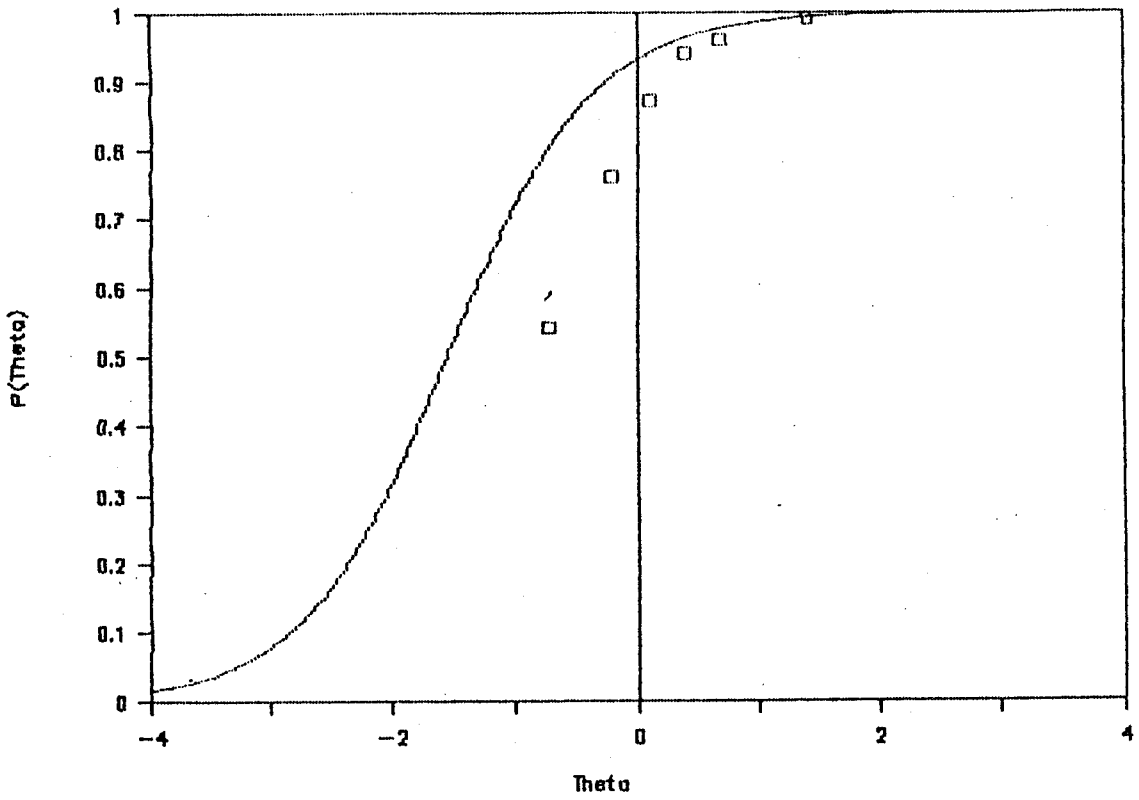
ITEM 45



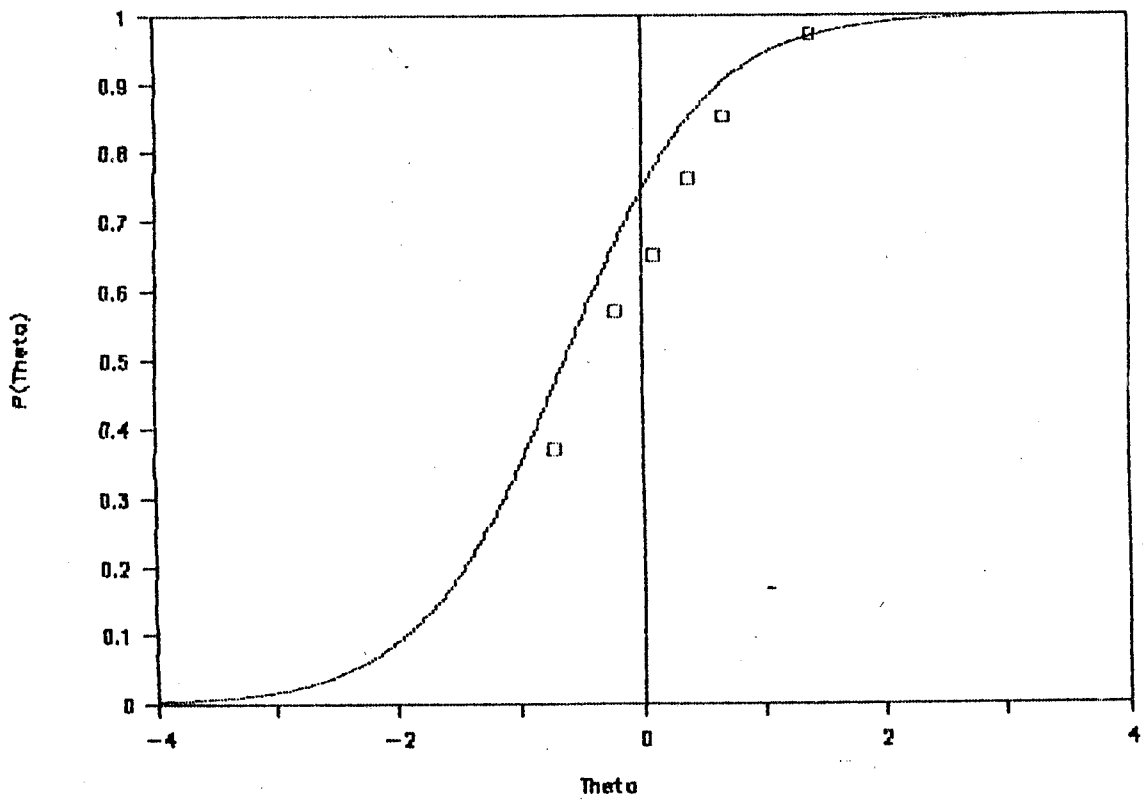
ITEM 46



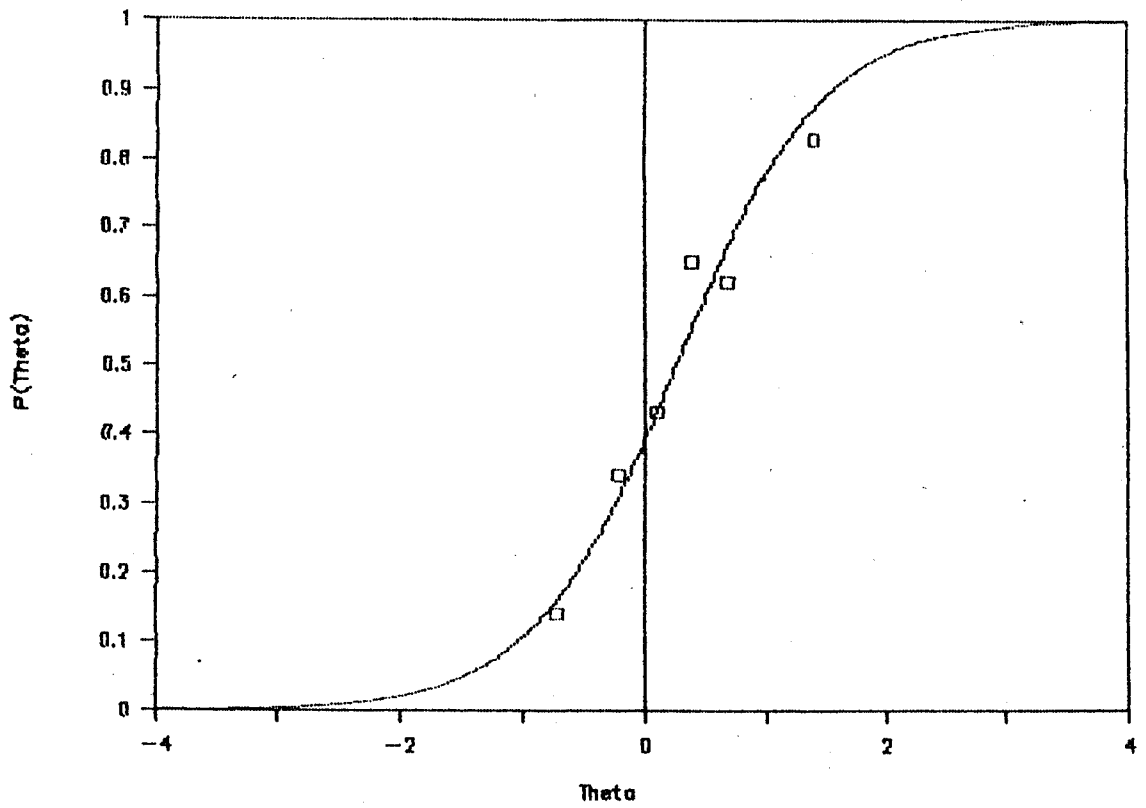
ITEM 47



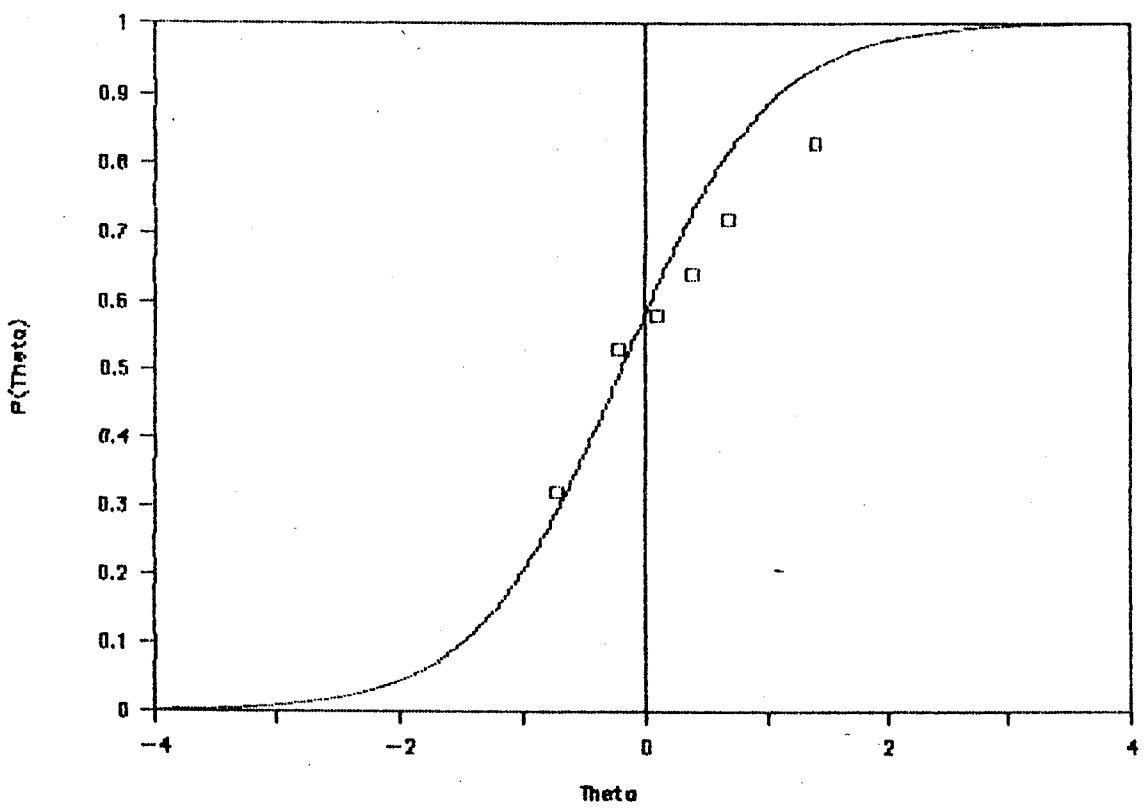
ITEM 48



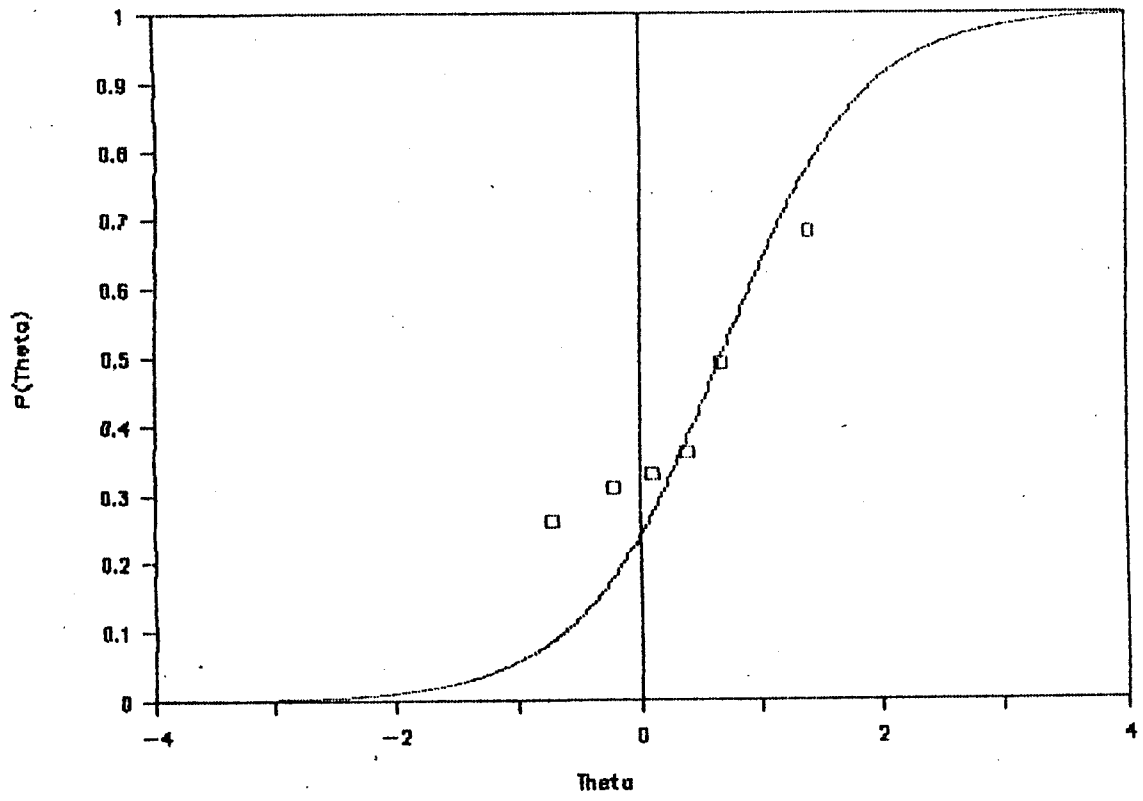
ITEM 49



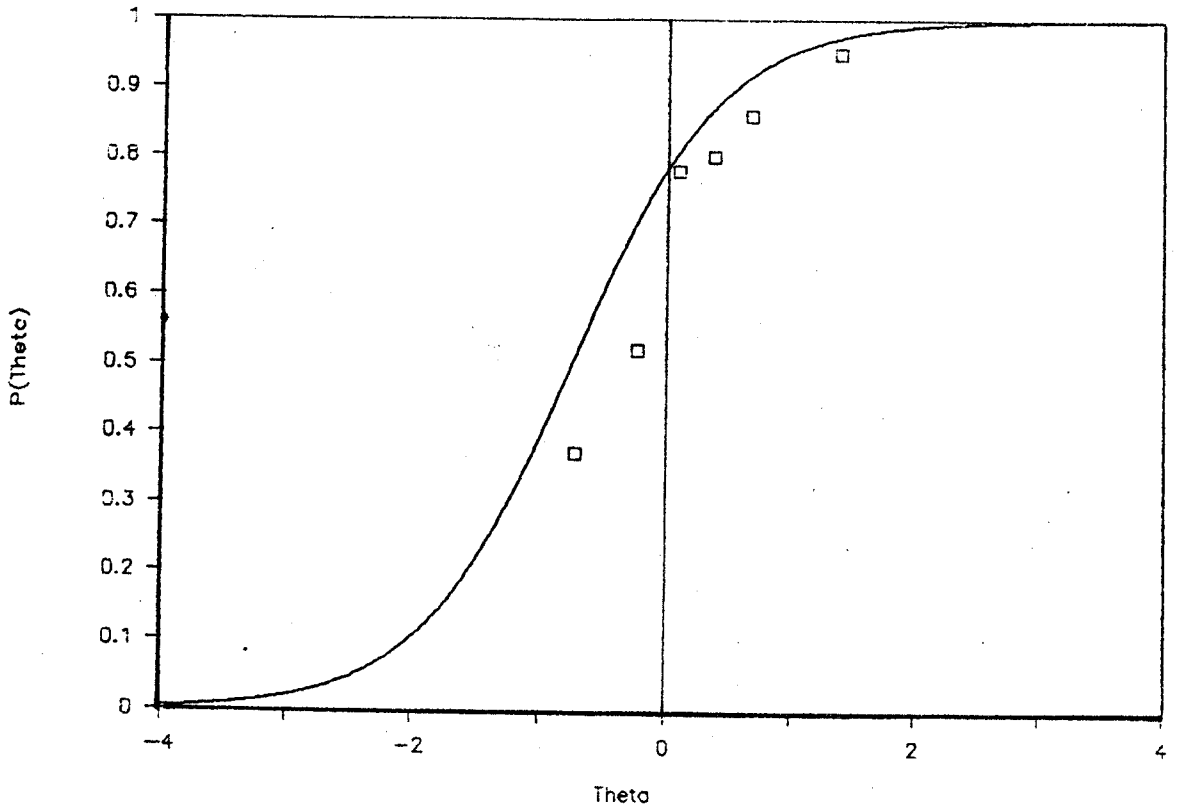
ITEM 50



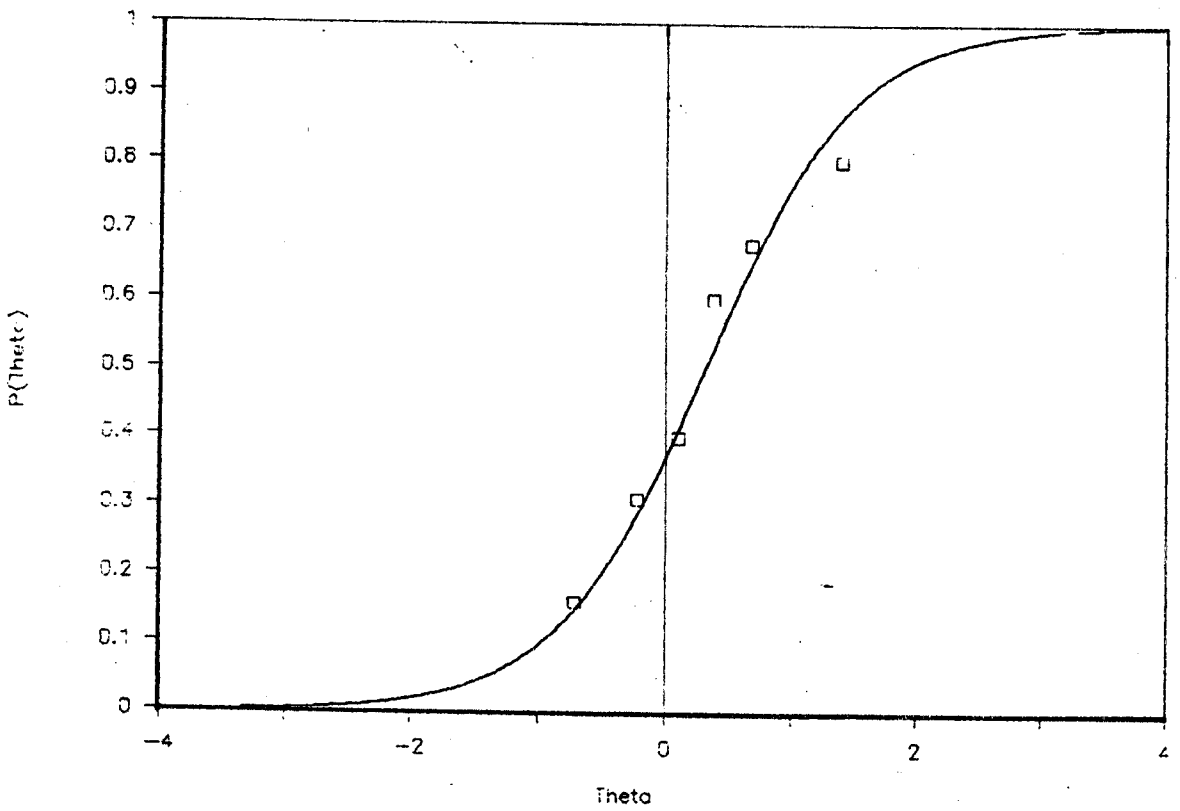
ITEM 51



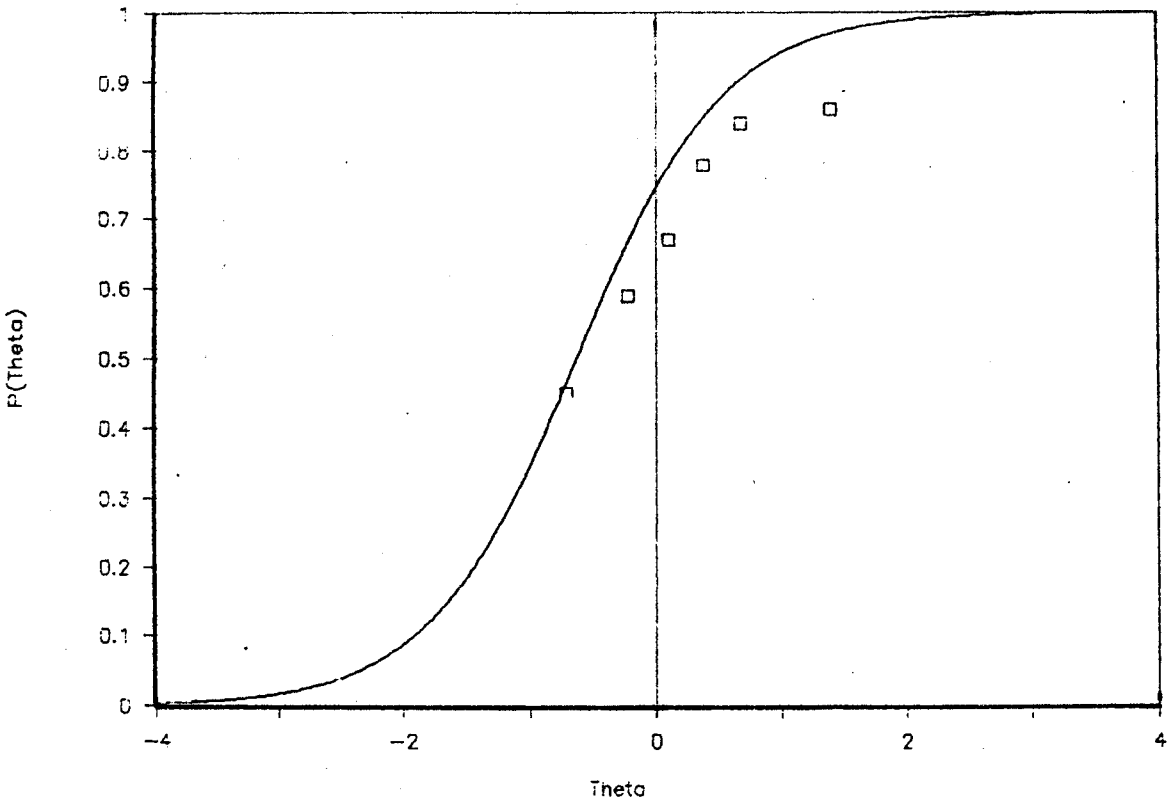
ITEM 52



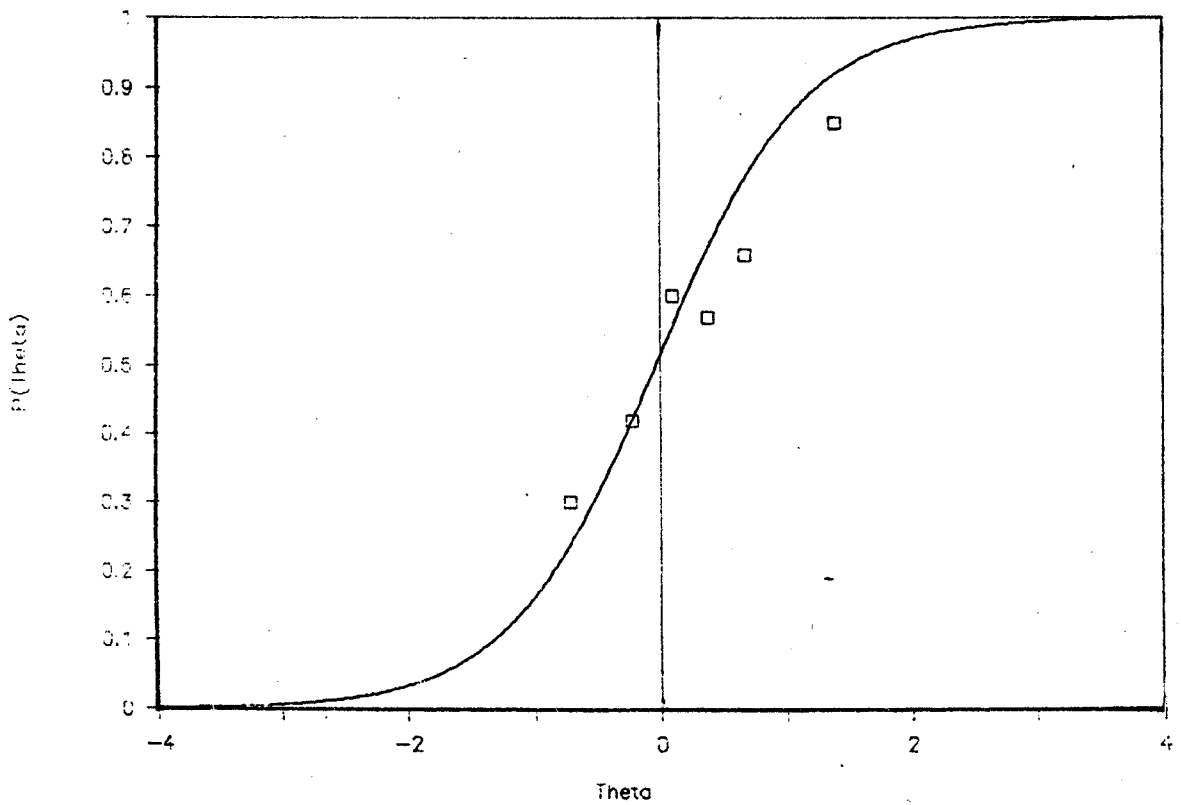
ITEM 53



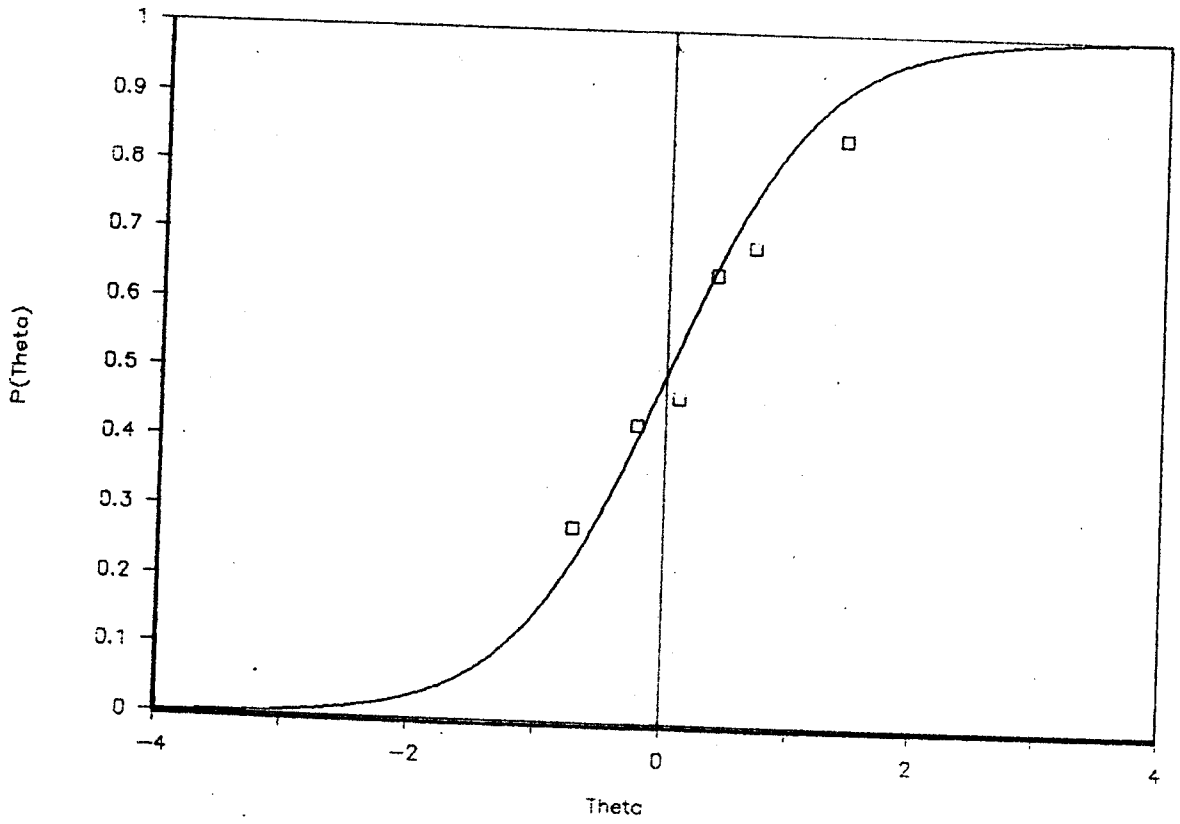
ITEM 54



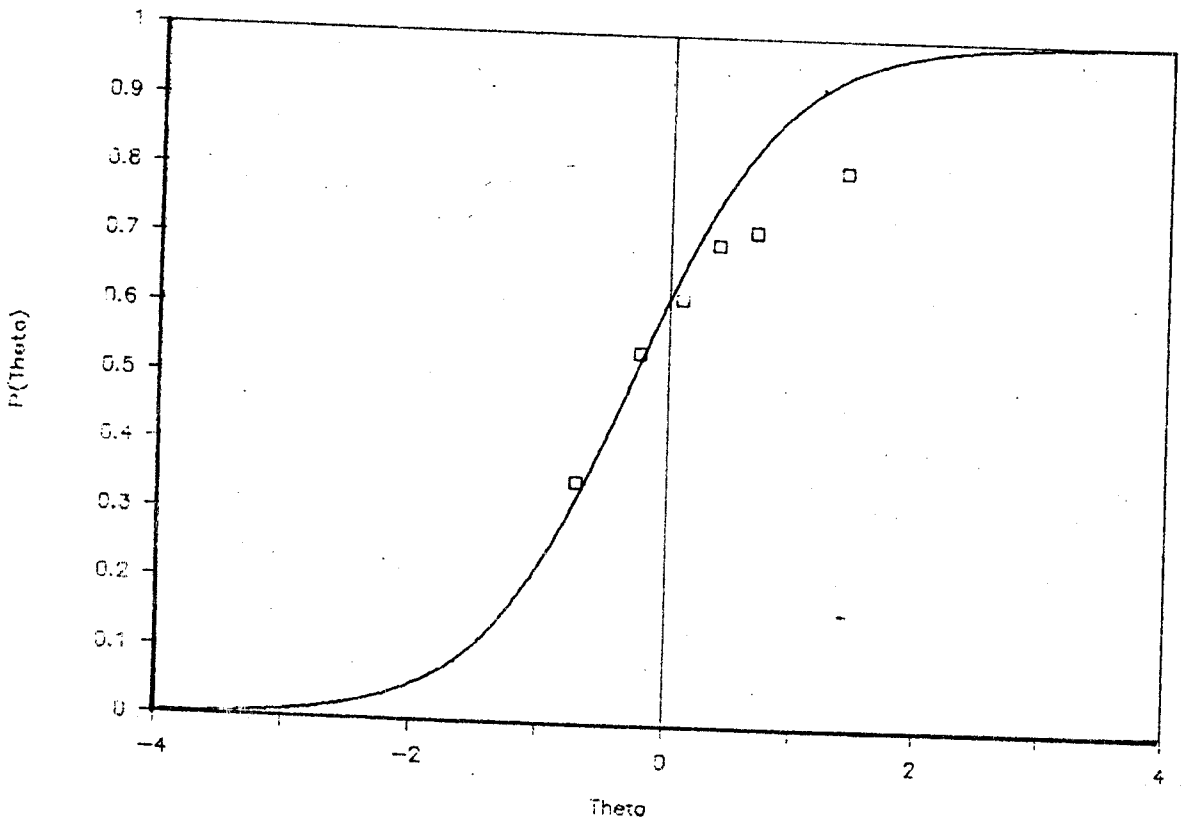
ITEM 55



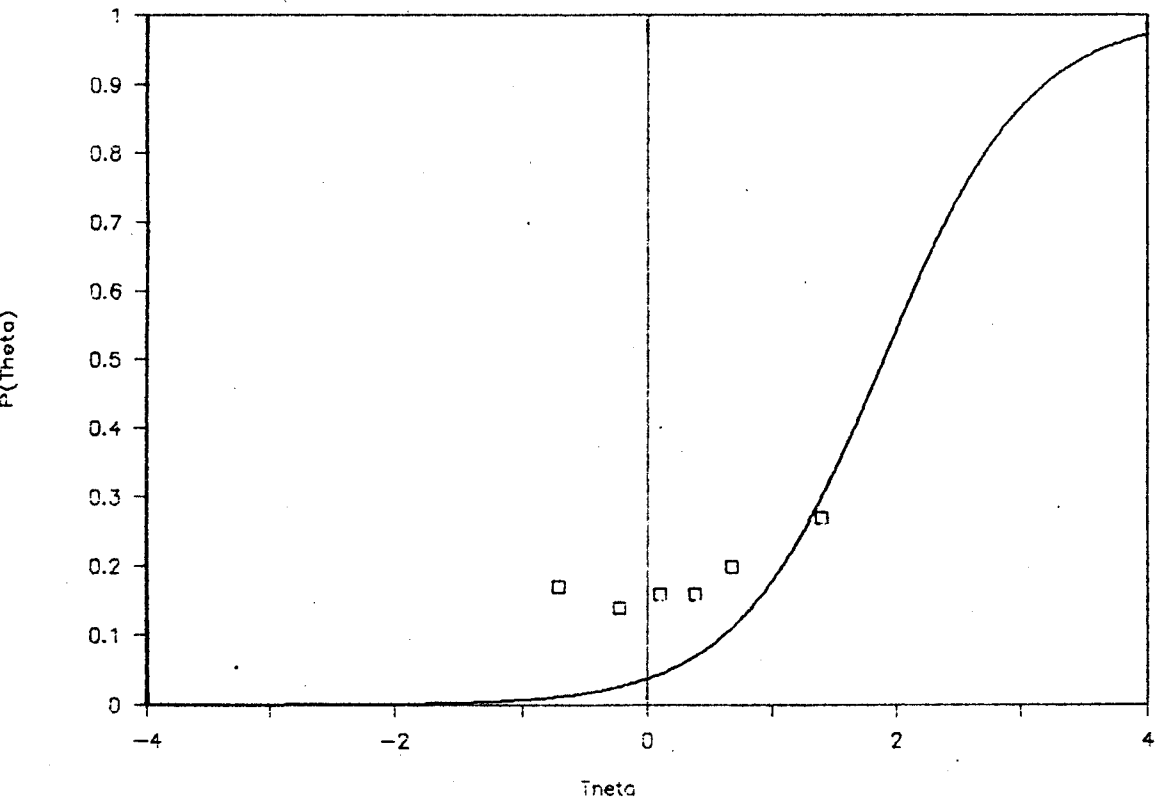
ITEM 56



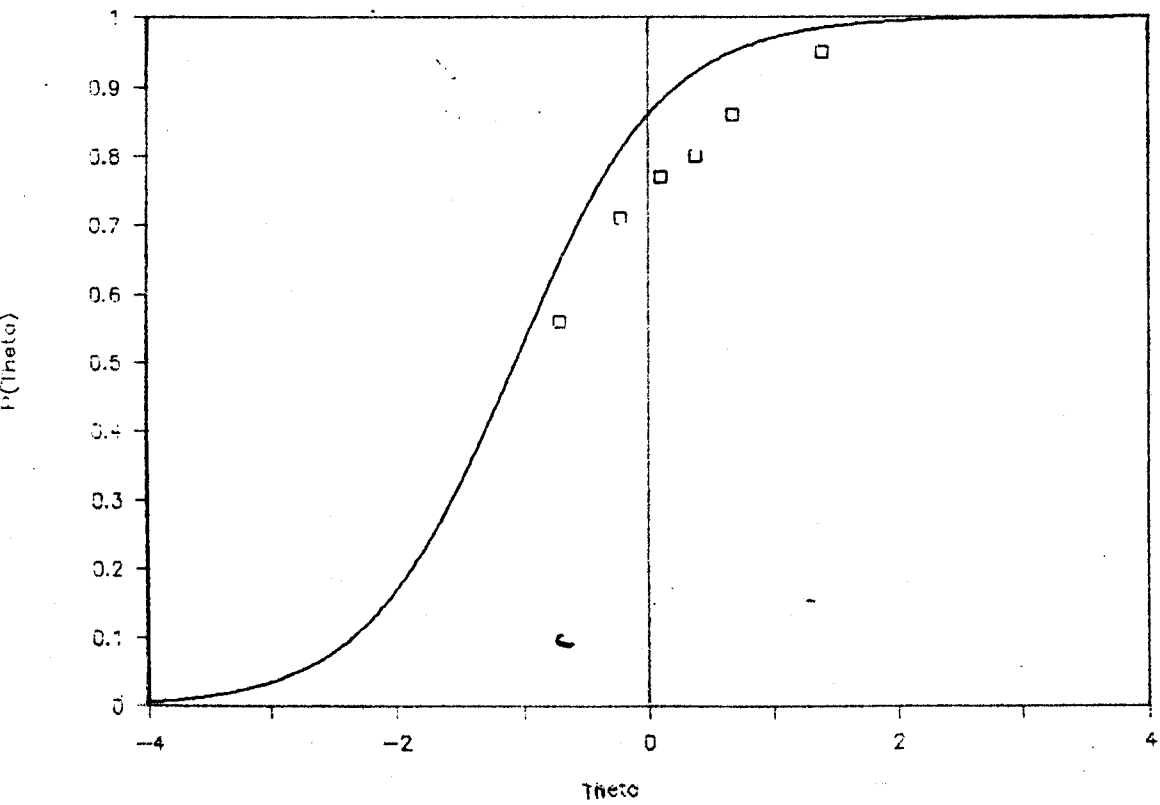
ITEM 57



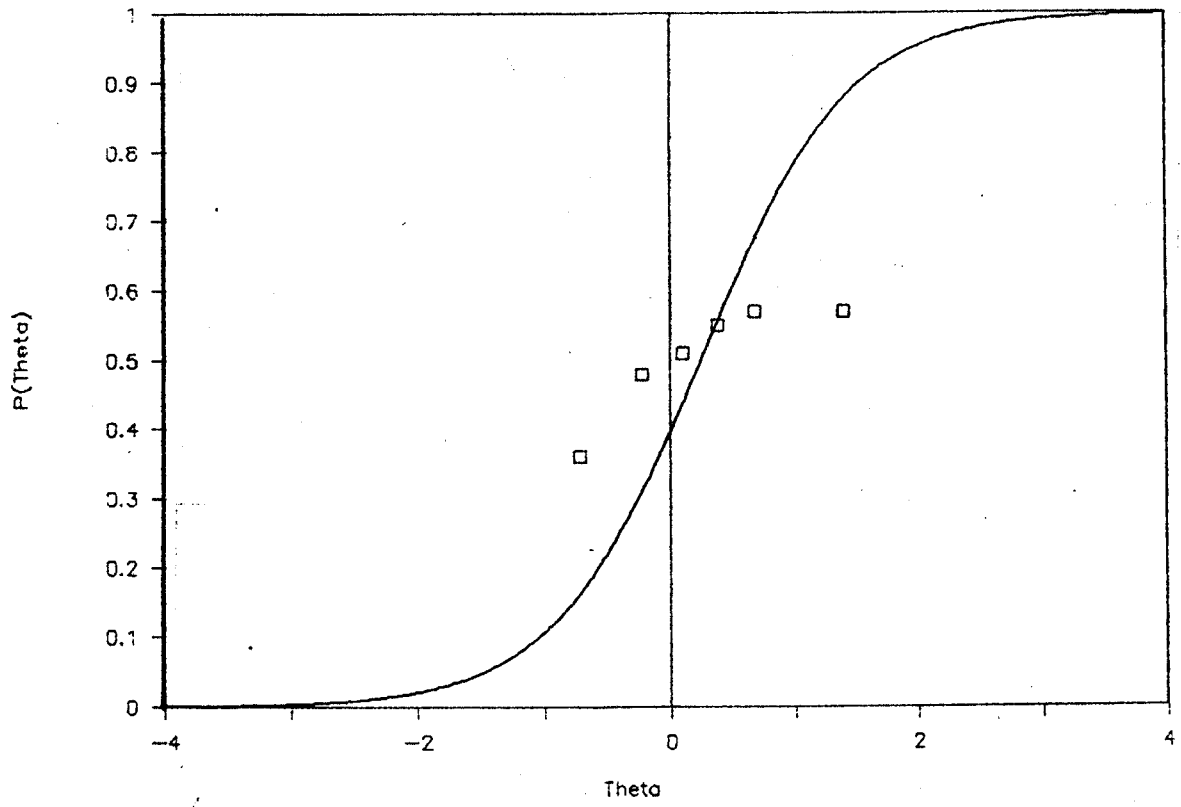
ITEM 58



ITEM 59



ITEM 60



ภาคผนวก ง

แสดงค่า p ทั้ง 6 ค่าในช่วงคะแนน 6 ช่วง
สำหรับแต่ละข้อ ($n=1485$)

ลำดับที่	ข้อที่	1st Group	2nd Group	3rd Group	4th Group	5th Group	6th Group
1	1	.74	.89	.94	.96	.98	1.00
2	2	.27	.47	.57	.70	.79	.90
3	3	.81	.92	.93	.94	.97	.98
4	4	.25	.32	.39	.37	.51	.71
5	5	.24	.29	.37	.45	.56	.72
6	6	.09	.18	.20	.19	.25	.44
7	7	.70	.74	.84	.88	.90	.97
8	8	.59	.76	.75	.83	.93	.92
9	9	.70	.82	.85	.84	.88	.96
10	10	.35	.35	.44	.57	.61	.79
11	11	.43	.63	.64	.77	.80	.92
12	12	.67	.77	.82	.80	.88	.94
13	13	.19	.17	.20	.31	.28	.48
14	15	.44	.56	.57	.56	.72	.72
15	16	.43	.53	.58	.59	.62	.76
16	17	.87	.93	.95	.94	.97	.98
17	18	.67	.76	.80	.85	.91	.94

ลำดับที่	ข้อที่	1st Group	2nd Group	3rd Group	4th Group	5th Group	6th Group
18	19	.54	.66	.68	.77	.84	.90
19	20	.34	.39	.45	.56	.55	.82
20	21	.42	.48	.50	.55	.65	.84
21	22	.18	.20	.23	.23	.25	.43
22	23	.33	.41	.57	.66	.76	.88
23	24	.44	.58	.64	.68	.72	.88
24	25	.18	.20	.23	.36	.27	.39
25	26	.23	.38	.56	.55	.69	.79
26	27	.20	.28	.40	.37	.59	.68
27	28	.18	.23	.25	.29	.37	.58
28	29	.47	.61	.71	.81	.87	.93
29	30	.49	.55	.60	.56	.62	.69
30	31	.33	.48	.61	.70	.83	.92
31	32	.50	.51	.58	.66	.71	.82
32	33	.33	.27	.34	.40	.44	.66
33	34	.25	.26	.32	.31	.34	.41
34	35	.29	.31	.40	.46	.47	.51

ลำดับที่	ข้อที่	1st Group	2nd Group	3rd Group	4th Group	5th Group	6th Group
35	36	.21	.28	.32	.36	.41	.62
36	37	.15	.16	.29	.36	.45	.69
37	38	.17	.20	.23	.30	.28	.58
38	39	.10	.07	.17	.20	.23	.46
39	40	.29	.43	.60	.60	.76	.81
40	41	.09	.14	.17	.21	.22	.30
41	42	.18	.18	.24	.28	.34	.53
42	43	.22	.30	.41	.50	.54	.80
43	44	.29	.37	.41	.45	.59	.77
44	45	.23	.49	.60	.74	.83	.94
45	46	.23	.38	.35	.46	.57	.62
46	47	.54	.76	.87	.94	.96	.99
47	48	.37	.57	.65	.76	.85	.97
48	49	.14	.34	.43	.65	.62	.83
49	50	.32	.53	.58	.64	.72	.83
50	51	.26	.31	.33	.36	.49	.68
51	52	.37	.52	.78	.80	.86	.95

ลำดับที่	ข้อที่	1st Group	2nd Group	3rd Group	4th Group	5th Group	6th Group
52	53	.16	.31	.40	.60	.68	.80
53	54	.45	.59	.67	.78	.84	.86
54	55	.30	.42	.60	.57	.66	.85
55	56	.28	.43	.47	.65	.69	.85
56	57	.35	.54	.62	.70	.72	.81
57	58	.17	.14	.16	.16	.20	.27
58	59	.56	.71	.77	.80	.86	.95
59	60	.36	.48	.51	.55	.57	.57
ช่วงคะแนน		1-24	25-28	29-32	33-35	36-39	40-58
ความสามารถเฉลี่ย		-.71	-.22	.10	.38	.68	1.40
จำนวนนักเรียนในกลุ่ม		227	261	282	220	229	266

ภาคผนวก จ

แสดงรายละเอียดผลการวิเคราะห์รายข้อ

ตามแบบดั้งเดิม (n=1502)

	Choice	Low				High
		N 1	N 2	N 3	N 4	N 5
	A	26 (.09)	15 (.04)	4 (.01)	3 (.01)	1 (.004)
	*B	227 (.76)	319 (.91)	277 (.95)	286 (.98)	268 (.99)
	C	27 (.09)	6 (.02)	3 (.03)	1 (.003)	1 (.004)
	D	17 (.06)	10 (.03)	3 (.01)	2 (.007)	0 (.000)
Base N	A	B	C	D	N _{total}	P ₊
1502	49	1375	44	32	32.25	.92
Item No.	N _A	N _B	N _C	N _D	P _{total}	r _{ct}
1	24.39	32.96	24.55	24.47	.99	.26

	Choice	low N 1	N 2	N 3	N 4	High N 5
	A	54 (.18)	40 (.12)	23 (.08)	20 (.07)	8 (.03)
	B	29 (.10)	22 (.06)	10 (.03)	11 (.04)	2 (.01)
	C	123 (.41)	112 (.32)	67 (.23)	39 (.13)	10 (.07)
	*D	91 (.31)	174 (.50)	194 (.66)	222 (.76)	242 (.90)
Base N	A	B	C	D	N _{total}	P ₊
1502	145	74	359	923	32.25	.61
Item No.	N _A	N _B	N _C	N _D	P _{total}	r _{ct}
2	28.02	26.97	28.01	34.99	.99	.37

	Choice	Low				High
		N 1	N 2	N 3	N 4	N 5
	A	24 (.08)	19 (.06)	8 (.03)	5 (.02)	3 (.01)
	B	14 (.05)	5 (.01)	2 (.01)	3 (.01)	2 (.01)
	*C	246 (.83)	319 (.92)	273 (.93)	282 (.97)	264 (.98)
	D	14 (.05)	5 (.01)	11 (.04)	1 (.003)	1 (.004)
Base N	A	B	C	D	N	P ₊
1502	59	26	1384	32	32.24	.92
Item No.	N _A	N _B	N _C	N _D	P _{total}	r _{ct}
3	26.76	26.92	32.72	25.88	.99	.18

	Choice	Low				High
		N 1	N 2	N 3	N 4	N 5
	A	79 (.27)	74 (.21)	64 (.22)	31 (.11)	10 (.04)
	B	37 (.12)	69 (.20)	59 (.20)	64 (.22)	43 (.16)
	C	105 (.35)	84 (.24)	65 (.22)	51 (.18)	28 (.10)
	*D	77 (.26)	121 (.35)	106 (.36)	146 (.50)	189 (.70)
Base N.	A	B	C	D	N	P ₊
1502	258	272	333	639	32.24	.43
Item No.	N _A	N _B	N _C	N _D	P _{total}	r _{ct}
4	28.82	32.54	29.18	35.09	1.0	.24

	Choice	Low				High
		N 1	N 2	N 3	N 4	N 5
	A	60 (.20)	43 (.12)	27 (.09)	9 (.03)	10 (.04)
	B	71 (.24)	64 (.18)	54 (.18)	43 (.15)	14 (.05)
	*C	70 (.24)	119 (.34)	111 (.38)	157 (.54)	196 (.73)
	D	97 (.33)	122 (.35)	101 (.35)	83 (.28)	49 (.18)
Base N	A	B	C	D	M	P ₊
1502	149	246	653	452	32.24	.43
Item No.	M _A	M _B	M _C	M _D	P _{total}	r _{ct}
5	27.14	29.36	35.48	30.80	.99	.30

	Choice	Low				High
		N 1	N 2	N 3	N 4	N 5
	A	171 (.58)	210 (.60)	179 (.61)	169 (.58)	126 (.47)
	B	26 (.09)	15 (.04)	18 (.06)	19 (.07)	11 (.04)
	C	69 (.23)	57 (.16)	40 (.14)	36 (.12)	18 (.07)
	*D	31 (.10)	66 (.19)	57 (.19)	68 (.23)	115 (.43)
Base N	A	B	C	D	M	P ₊
1502	855	89	220	337	32.25	.22
Item No.	M _A	M _B	M _C	M _D	P _{total}	r _{ct}
6	31.78	30.30	29.07	36.02	.99	.21

	Choice	Low				High
		N 1	N 2	N 3	N 4	N 5
	A	81 (.27)	61 (.18)	42 (.14)	28 (.10)	7 (.03)
	*B	208 (.70)	275 (.79)	251 (.85)	259 (.89)	251 (.97)
	C	4 (.01)	11 (.03)	0 (.000)	4 (.01)	2 (.01)
	D	4 (.01)	1 (.003)	1 (.003)	1 (.003)	0 (.000)
Base N	A	B	C	D	M	P ₊
1502	219	1254	21	7	32.26	.84
Item No.	M _A	M _B	M _C	M _D	P _{total}	r _{ct}
7	27.70	33.15	29.38	24.14	.99	.22

	Choice	Low				High
		N 1	N 2	N 3	N 4	N 5
	A	41 (.14)	21 (.06)	15 (.05)	7 (.02)	3 (.01)
	B	26 (.09)	23 (.07)	21 (.07)	10 (.03)	4 (.02)
	C	48 (.16)	34 (.10)	32 (.11)	9 (.03)	13 (.05)
	*D	183 (.61)	270 (.78)	224 (.77)	266 (.91)	250 (.93)
Base N	A	B	C	D	M	P ₊
1502	87	84	136	1193	32.24	.79
Item No.	M _A	M _B	M _C	M _D	P _{total}	r _{ct}
8	26.57	28.58	28.44	33.35	.99	.23

	Choice	Low					High
		N 1	N 2	N 3	N 4	N 5	
A	15	17	19	8	3		
	(.05)	(.05)	(.07)	(.03)	(.01)		
B	12	3	7	4	2		
	(.04)	(.01)	(.02)	(.01)	(.01)		
*C	215	295	244	255	256		
	(.72)	(.85)	(.83)	(.87)	(.95)		
D	56	33	24	25	9		
	(.19)	(.10)	(.08)	(.09)	(.03)		
Base N	A	B	C	D	N	P ₊	
1502	62	28	1265	147	32.24	.84	
Item No.	N _A	N _B	N _C	N _D	P _{total}	r _{it}	
9	29.06	27.89	32.91	28.65	1.0	.16	

	Choice	Low					High
		N 1	N 2	N 3	N 4	N 5	
A	99	141	75	68	33		
	(.34)	(.41)	(.26)	(.23)	(.12)		
*B	101	142	144	175	210		
	(.34)	(.41)	(.49)	(.60)	(.78)		
C	50	37	30	18	17		
	(.17)	(.11)	(.10)	(.06)	(.06)		
D	47	28	45	31	10		
	(.16)	(.08)	(.15)	(.11)	(.04)		
Base N	A	B	C	D	N	P ₊	
1502	416	772	152	161	32.25	.51	
Item No.	N _A	N _B	N _C	N _D	P _{total}	r _{it}	
10	29.79	34.65	29.29	29.88	.99	.25	

	Choice	Low					High
		N 1	N 2	N 3	N 4	N 5	
*A	135	223	209	232	249		
	(.45)	(.64)	(.71)	(.80)	(.92)		
B	32	28	16	15	9		
	(.11)	(.08)	(.05)	(.05)	(.03)		
C	102	81	55	41	11		
	(.34)	(.23)	(.19)	(.14)	(.04)		
D	29	15	14	4	1		
	(.10)	(.04)	(.05)	(.01)	(.004)		
Base N	A	B	C	D	N	P ₊	
1502	1048	100	290	63	32.25	.70	
Item No.	N _A	N _B	N _C	N _D	P _{total}	r _{it}	
11	34.07	29.12	28.14	25.79	.99	.29	

	Choice	Low					High
		N 1	N 2	N 3	N 4	N 5	
*A	201	274	244	248	254		
	(.67)	(.79)	(.83)	(.85)	(.94)		
B	40	26	20	11	6		
	(.13)	(.08)	(.07)	(.04)	(.02)		
C	21	19	14	13	3		
	(.07)	(.06)	(.05)	(.05)	(.01)		
D	36	29	16	20	7		
	(.12)	(.08)	(.05)	(.07)	(.03)		
Base N	A	B	C	D	N	P ₊	
1502	1221	103	70	108	32.24	.81	
Item No.	N _A	N _B	N _C	N _D	P _{total}	r _{it}	
12	33.12	27.97	28.89	28.56	1.0	.19	

Choice	Low					High
	N 1	N 2	N 3	N 4	N 5	
A	171 (.58)	215 (.62)	164 (.56)	172 (.59)	126 (.47)	
B	60 (.20)	64 (.18)	49 (.17)	29 (.10)	16 (.06)	
C	12 (.04)	9 (.03)	4 (.01)	2 (.01)	1 (.004)	
*D	54 (.18)	60 (.17)	77 (.26)	89 (.31)	127 (.47)	
Base N	A	B	C	D	N	P ₊
1502	848	218	28	407	32.25	.27
Item No.	N _A	N _B	N _C	N _D	P _{total}	r _{ct}
13	31.67	29.42	26.96	35.34	.99	.18

Choice	Low					High
	N 1	N 2	N 3	N 4	N 5	
*A	69 (.23)	86 (.25)	81 (.28)	77 (.26)	53 (.20)	
B	31 (.10)	19 (.06)	8 (.03)	7 (.02)	3 (.01)	
C	185 (.62)	233 (.67)	196 (.67)	205 (.70)	212 (.79)	
D	13 (.04)	10 (.03)	9 (.03)	3 (.01)	2 (.01)	
Base N	A	B	C	D	N	P ₊
1502	366	68	1031	37	32.24	.24
Item No.	N _A	N _B	N _C	N _D	P _{total}	r _{ct}
14	31.87	26.29	32.90	28.54	1.0	-.09

Choice	Low					High
	N 1	N 2	N 3	N 4	N 5	
A	90 (.30)	99 (.28)	67 (.23)	47 (.16)	36 (.13)	
B	29 (.01)	28 (.08)	25 (.09)	20 (.07)	14 (.05)	
C	133 (.45)	191 (.55)	170 (.58)	201 (.69)	195 (.72)	
D	46 (.15)	30 (.09)	32 (.11)	24 (.08)	25 (.09)	
Base N	A	B	C	D	N	P ₊
1502	339	116	890	157	32.24	.59
Item No.	N _A	N _B	N _C	N _D	P _{total}	r _{ct}
15	29.78	30.77	33.57	31.11	1.0	.14

Choice	Low					High
	N 1	N 2	N 3	N 4	N 5	
A	41 (.14)	20 (.06)	16 (.05)	22 (.08)	5 (.02)	
B	61 (.21)	60 (.17)	49 (.17)	48 (.16)	27 (.10)	
C	66 (.22)	81 (.23)	52 (.18)	46 (.16)	34 (.13)	
*D	130 (.44)	187 (.54)	177 (.60)	176 (.60)	204 (.75)	
Base N	A	B	C	D	N	P ₊
1502	104	245	279	874	32.24	.58
Item No.	N _A	N _B	N _C	N _D	P _{total}	r _{ct}
16	28.56	30.59	30.31	33.76	1.0	.19

Choice	Low					High	
	N 1	N 2	N 3	N 4	N 5		
A	23 (.08)	12 (.03)	11 (.04)	5 (.02)	4 (.02)		
*B	259 (.87)	330 (.95)	274 (.93)	284 (.97)	262 (.97)		
C	9 (.03)	4 (.01)	9 (.03)	3 (.01)	3 (.01)		
D	7 (.02)	2 (.01)	0 (.000)	0 (.000)	1 (.004)		
Base N	A	B	C	D	N	P ₊	
2	55	1409	28	10	32.24	.94	
Item No.	N _A	N _B	N _C	N _D	P _{total}	r _{ct}	
17	27.55	32.55	29.50	23.10	1.0	.13	

Choice	Low					High	
	N 1	N 2	N 3	N 4	N 5		
A	67 (.23)	63 (.18)	37 (.13)	21 (.07)	10 (.04)		
*B	196 (.66)	269 (.77)	245 (.83)	264 (.90)	255 (.94)		
C	20 (.07)	12 (.03)	7 (.02)	4 (.01)	4 (.02)		
D	15 (.05)	4 (.01)	5 (.02)	3 (.01)	1 (.004)		
Base N	A	B	C	D	N	P ₊	
1502	198	1229	47	28	32.24	.82	
Item No.	N _A	N _B	N _C	N _D	P _{total}	r _{ct}	
18	28.29	33.23	26.96	25.89	1.0	.20	

Choice	Low					High	
	N 1	N 2	N 3	N 4	N 5		
A	76 (.26)	84 (.24)	57 (.20)	32 (.11)	16 (.06)		
B	33 (.11)	11 (.03)	8 (.03)	3 (.01)	7 (.03)		
*C	161 (.54)	234 (.67)	212 (.72)	241 (.83)	241 (.89)		
D	27 (.09)	19 (.06)	16 (.06)	16 (.06)	6 (.02)		
Base N	A	B	C	D	N	P ₊	
	265	62	1089	84	32.25	.73	
Item No.	N _A	N _B	N _C	N _D	P _{total}	r _{ct}	
9	28.84	26.24	33.68	28.93	.99	.25	

Choice	Low					High	
	N 1	N 2	N 3	N 4	N 5		
*A	97 (.33)	147 (.42)	150 (.51)	165 (.57)	219 (.81)		
B	139 (.47)	141 (.41)	100 (.34)	85 (.29)	35 (.13)		
C	32 (.11)	28 (.08)	14 (.05)	15 (.05)	8 (.03)		
D	30 (.10)	32 (.09)	30 (.10)	24 (.08)	8 (.03)		
Base N	A	B	C	D	N	P ₊	
1502	778	500	97	124	32.23	.52	
Item No.	N _A	N _B	N _C	N _D	P _{total}	r _{ct}	
20	34.72	29.54	28.75	30.24	.99	.27	

Choice	Low					High
	N 1	N 2	N 3	N 4	N 5	
*A	120 (.40)	176 (.51)	156 (.53)	160 (.62)	225 (.83)	
B	52 (.17)	70 (.20)	57 (.20)	36 (.12)	15 (.06)	
C	40 (.13)	21 (.06)	12 (.04)	12 (.04)	4 (.02)	
D	86 (.29)	81 (.23)	68 (.23)	64 (.22)	26 (.10)	
Base N	A	B	C	D	N	P ₊
1502	857	230	89	325	32.24	.57
Item No.	N _A	N _B	N _C	N _D	P _{total}	r _{ct}
21	34.25	30.08	26.67	30.00	.99	.24

Choice	Low					High
	N 1	N 2	N 3	N 4	N 5	
A	83 (.28)	112 (.32)	89 (.30)	92 (.32)	58 (.22)	
B	92 (.31)	110 (.32)	104 (.35)	98 (.34)	86 (.32)	
*C	53 (.18)	71 (.21)	70 (.24)	73 (.25)	115 (.43)	
D	70 (.24)	54 (.16)	31 (.11)	27 (.09)	11 (.04)	
Base N	A	B	C	D	N	P ₊
1502	434	490	382	193	32.24	.25
Item No.	N _A	N _B	N _C	N _D	P _{total}	r _{ct}
22	31.57	32.28	35.10	27.98	.99	.15

Choice	Low					High
	N 1	N 2	N 3	N 4	N 5	
A	135 (.45)	150 (.43)	97 (.33)	56 (.23)	24 (.09)	
*B	98 (.33)	165 (.47)	178 (.61)	219 (.75)	237 (.88)	
C	35 (.12)	20 (.06)	10 (.03)	4 (.01)	6 (.02)	
D	30 (.10)	13 (.04)	9 (.03)	2 (.01)	3 (.01)	
Base N	A	B	C	D	N	P ₊
1502	472	897	75	57	32.24	.60
Item No.	N _A	N _B	N _C	N _D	P _{total}	r _{ct}
23	28.84	34.88	26.99	25.81	.99	.34

Choice	Low					High
	N 1	N 2	N 3	N 4	N 5	
*A	132 (.44)	209 (.60)	204 (.70)	203 (.70)	235 (.87)	
B	59 (.20)	50 (.14)	28 (.10)	26 (.09)	12 (.04)	
C	60 (.20)	52 (.15)	37 (.13)	41 (.14)	14 (.05)	
D	47 (.16)	37 (.11)	24 (.08)	22 (.08)	9 (.03)	
Base N	A	B	C	D	N	P ₊
1502	983	175	204	139	32.24	.65
Item No.	N _A	N _B	N _C	N _D	P _{total}	r _{ct}
24	34.05	28.45	29.06	28.90	.99	.26

	Choice	Low					High
		N 1	N 2	N 3	N 4	N 5	
*A	51	76	84	95	103		
	(.17)	(.22)	(.29)	(.33)	(.38)		
B	57	34	28	25	26		
	(.19)	(.10)	(.10)	(.09)	(.10)		
C	128	172	140	120	92		
	(.43)	(.49)	(.48)	(.41)	(.34)		
D	61	66	42	51	49		
	(.21)	(.19)	(.14)	(.18)	(.18)		
Base N	A	B	C	D	N	P ₊	
1502	409	170	652	269	32.25	.27	
Item No.	N _A	N _B	N _C	N _D	P _{total}	r _{ct}	
25	34.54	30.40	31.46	31.85	.99	.11	

	Choice	Low					High
		N 1	N 2	N 3	N 4	N 5	
A	90	82	69	45	24		
	(.30)	(.24)	(.24)	(.16)	(.09)		
B	45	44	26	21	13		
	(.16)	(.13)	(.09)	(.07)	(.05)		
*C	79	148	165	197	209		
	(.27)	(.43)	(.56)	(.68)	(.77)		
D	84	73	34	28	24		
	(.28)	(.21)	(.12)	(.10)	(.09)		
Base N	A	B	C	D	N	P ₊	
1502	310	149	798	243	32.25	.53	
Item No.	N _A	N _B	N _C	N _D	P _{total}	r _{ct}	
26	29.42	29.00	35.09	28.49	.99	.32	

	Choice	Low					High
		N 1	N 2	N 3	N 4	N 5	
A	51	52	27	28	18		
	(.17)	(.15)	(.09)	(.10)	(.07)		
B	101	128	107	78	50		
	(.34)	(.37)	(.36)	(.27)	(.19)		
C	82	52	39	38	15		
	(.28)	(.15)	(.13)	(.13)	(.06)		
*D	63	116	121	148	187		
	(.21)	(.33)	(.41)	(.51)	(.69)		
Base N	A	B	C	D	N	P ₊	
1502	176	464	226	635	32.26	.42	
Item No.	N _A	N _B	N _C	N _D	P _{total}	r _{ct}	
27	29.31	30.72	29.05	35.34	.99	.26	

	Choice	Low					High
		N 1	N 2	N 3	N 4	N 5	
A	67	88	75	75	57		
	(.23)	(.25)	(.26)	(.26)	(.21)		
B	118	134	100	78	41		
	(.40)	(.39)	(.34)	(.27)	(.15)		
C	58	44	40	30	22		
	(.20)	(.13)	(.14)	(.10)	(.08)		
*D	54	80	79	108	150		
	(.18)	(.23)	(.27)	(.37)	(.56)		
Base N	A	B	C	D	N	P ₊	
1502	362	471	194	471	32.26	.31	
Item No.	N _A	N _B	N _C	N _D	P _{total}	r _{ct}	
28	32.10	29.80	30.05	35.76	.99	.24	

	Choice	Low					High
		N 1	N 2	N 3	N 4	N 5	
	A	42 (.14)	43 (.12)	24 (.08)	14 (.05)	10 (.04)	
	*B	143 (.48)	225 (.65)	224 (.77)	251 (.86)	250 (.93)	
	C	54 (.18)	43 (.12)	22 (.08)	20 (.07)	6 (.02)	
	D	58 (.20)	37 (.11)	23 (.08)	7 (.02)	4 (.02)	
Base N	A	B	C	D	M	P+	
1502	133	1093	145	129	32.26	.73	
Item No.	N A	N B	N C	N D	P total	r it	
29	28.48	34.04	27.63	26.22	.99	.31	

	Choice	Low					High
		N 1	N 2	N 3	N 4	N 5	
	A	42 (.14)	38 (.11)	25 (.09)	37 (.13)	19 (.07)	
	*B	144 (.49)	192 (.55)	172 (.59)	178 (.61)	190 (.71)	
	C	72 (.24)	73 (.21)	65 (.22)	44 (.15)	34 (.13)	
	D	39 (.13)	44 (.13)	32 (.11)	33 (.11)	26 (.10)	
Base N	A	B	C	D	M	P+	
1502	161	876	288	174	32.25	.58	
Item No.	N A	N B	N C	N D	P total	r it	
30	30.94	33.29	30.38	31.33	.99	.07	

	Choice	Low					High
		N 1	N 2	N 3	N 4	N 5	
	*A	104 (.35)	192 (.55)	189 (.65)	232 (.80)	247 (.92)	
	B	49 (.17)	57 (.16)	42 (.14)	29 (.10)	12 (.04)	
	C	46 (.16)	34 (.10)	25 (.09)	13 (.05)	5 (.02)	
	D	98 (.33)	65 (.19)	37 (.13)	17 (.06)	6 (.02)	
Base N	A	B	C	D	M	P+	
1502	964	189	123	223	32.25	.64	
Item No.	N A	N B	N C	N D	P total	r it	
31	34.69	29.58	27.49	26.56	.99	.36	

	Choice	Low					High
		N 1	N 2	N 3	N 4	N 5	
	A	55 (.19)	91 (.26)	70 (.24)	55 (.19)	35 (.13)	
	B	74 (.25)	51 (.15)	27 (.09)	21 (.07)	9 (.03)	
	*C	144 (.48)	186 (.54)	182 (.62)	207 (.71)	219 (.81)	
	D	25 (.08)	19 (.06)	14 (.05)	9 (.03)	7 (.03)	
Base N	A	B	C	D	M	P+	
1502	306	182	938	74	32.25	.62	
Item No.	N A	N B	N C	N D	P total	r it	
32	31.21	27.21	33.84	28.73	.99	.20	

Choice	Low					High	
	N 1	N 2	N 3	N 4	N 5		
A	97 (.33)	136 (.39)	97 (.33)	101 (.35)	67 (.25)		
B	28 (.09)	28 (.08)	23 (.08)	26 (.09)	10 (.04)		
C	81 (.27)	78 (.22)	64 (.22)	37 (.13)	16 (.06)		
*D	92 (.31)	106 (.31)	110 (.37)	127 (.44)	177 (.66)		
Base N	A	B	C	D	M	P+	
1502	498	115	276	612	32.24	.41	
Item No.	N A	N B	N C	N D	P total	r it	
33	31.51	30.54	28.91	34.66	.99	.19	

Choice	Low					High	
	N 1	N 2	N 3	N 4	N 5		
A	55 (.19)	51 (.15)	30 (.10)	27 (.09)	19 (.07)		
*B	71 (.24)	103 (.30)	88 (.30)	98 (.34)	113 (.42)		
C	94 (.32)	95 (.28)	94 (.32)	88 (.30)	72 (.27)		
D	76 (.26)	97 (.28)	81 (.28)	76 (.26)	66 (.24)		
Base N	A	B	C	D	M	P+	
1502	182	473	443	396	32.26	.31	
Item No.	N A	N B	N C	N D	P total	r it	
34	29.68	33.87	31.86	31.97	.99	.07	

Choice	Low					High	
	N 1	N 2	N 3	N 4	N 5		
A	106 (.36)	112 (.32)	71 (.24)	58 (.20)	22 (.08)		
*B	91 (.31)	115 (.33)	128 (.44)	131 (.45)	142 (.53)		
C	76 (.26)	94 (.27)	70 (.24)	68 (.24)	49 (.18)		
D	23 (.08)	25 (.07)	23 (.08)	32 (.11)	57 (.21)		
Base N	A	B	C	D	M	P+	
1502	369	607	357	160	32.26	.40	
Item No.	N A	N B	N C	N D	P total	r it	
35	28.95	33.98	31.25	35.58	.99	.11	

Choice	Low					High	
	N 1	N 2	N 3	N 4	N 5		
A	92 (.31)	106 (.31)	74 (.25)	65 (.22)	39 (.15)		
B	59 (.20)	58 (.17)	37 (.13)	39 (.13)	17 (.06)		
C	81 (.27)	83 (.24)	82 (.28)	66 (.23)	46 (.17)		
*D	66 (.22)	101 (.29)	98 (.34)	122 (.42)	166 (.62)		
Base N	A	B	C	D	M	P+	
1502	376	210	358	553	32.23	.37	
Item No.	N A	N B	N C	N D	P total	r it	
36	30.23	29.67	31.11	35.29	.99	.24	

Choice	Low					High
	N 1	N 2	N 3	N 4	N 5	
*A	49 (.17)	69 (.20)	102 (.35)	122 (.42)	185 (.69)	
B	101 (.34)	115 (.33)	75 (.26)	59 (.20)	45 (.17)	
C	91 (.31)	104 (.30)	73 (.25)	62 (.21)	19 (.07)	
D	55 (.19)	58 (.17)	43 (.15)	47 (.16)	20 (.07)	
Base N	A	B	C	D	N	P ₊
1502	527	395	349	223	32.25	.35
Item No.	N _A	N _B	N _C	N _D	P _{total}	r _{ct}
37	36.44	30.21	29.40	30.41	.99	.37

Choice	Low					High
	N 1	N 2	N 3	N 4	N 5	
A	25 (.08)	16 (.05)	14 (.05)	17 (.06)	5 (.02)	
B	140 (.47)	143 (.41)	100 (.34)	71 (.24)	27 (.10)	
*C	46 (.17)	74 (.21)	82 (.28)	80 (.27)	156 (.58)	
D	86 (.29)	113 (.33)	98 (.33)	124 (.43)	82 (.30)	
Base N	A	B	C	D	N	P ₊
1502	77	481	438	503	32.25	.29
Item No.	N _A	N _B	N _C	N _D	P _{total}	r _{ct}
38	29.41	28.79	36.07	32.68	.99	.24

Choice	Low					High
	N 1	N 2	N 3	N 4	N 5	
A	207 (.70)	252 (.73)	166 (.57)	141 (.48)	82 (.31)	
B	38 (.13)	43 (.12)	61 (.21)	69 (.24)	58 (.22)	
*C	28 (.10)	35 (.10)	58 (.20)	67 (.23)	122 (.46)	
D	23 (.08)	16 (.05)	8 (.03)	15 (.05)	4 (.02)	
Base N	A	B	C	D	N	P ₊
1502	848	269	310	66	32.23	.21
Item No.	N _A	N _B	N _C	N _D	P _{total}	r _{ct}
39	30.18	33.67	37.24	29.26	.99	.26

Choice	Low					High
	N 1	N 2	N 3	N 4	N 5	
A	93 (.31)	56 (.16)	28 (.10)	26 (.09)	6 (.02)	
B	71 (.24)	92 (.27)	57 (.19)	47 (.16)	24 (.09)	
*C	90 (.30)	170 (.49)	183 (.62)	204 (.70)	222 (.82)	
D	44 (.15)	29 (.08)	26 (.09)	15 (.05)	18 (.07)	
Base N	A	B	C	D	N	P ₊
1502	209	291	869	132	32.25	.58
Item No.	N _A	N _B	N _C	N _D	P _{total}	r _{ct}
40	26.45	29.91	34.80	29.80	.99	.32

Choice	Low					High
	N 1	N 2	N 3	N 4	N 5	
A	48 (.16)	26 (.08)	16 (.06)	5 (.02)	1 (.004)	
B	96 (.32)	97 (.28)	83 (.28)	84 (.29)	54 (.20)	
C	122 (.41)	172 (.50)	137 (.47)	139 (.48)	132 (.49)	
*D	31 (.10)	50 (.15)	56 (.19)	63 (.22)	33 (.31)	

Base N	A	B	C	D	N	P+
1502	96	414	702	283	32.26	.19
Item No.	N A	N B	N C	N D	P total	r ct
41	25.29	31.13	32.68	35.24	.99	.13

Choice	Low					High
	N 1	N 2	N 3	N 4	N 5	
A	47 (.16)	80 (.23)	61 (.21)	66 (.23)	47 (.18)	
B	88 (.29)	107 (.31)	74 (.25)	61 (.21)	44 (.16)	
*C	50 (.17)	78 (.23)	77 (.26)	92 (.32)	145 (.54)	
D	113 (.38)	81 (.23)	81 (.28)	73 (.25)	33 (.12)	

Base N	A	B	C	D	N	P+
1502	301	374	442	381	31.24	.29
Item No.	N A	N B	N C	N D	P total	r ct
42	32.46	30.64	35.67	29.67	.99	.22

Choice	Low					High
	N 1	N 2	N 3	N 4	N 5	
*A	70 (.24)	144 (.33)	142 (.49)	156 (.54)	213 (.79)	
B	95 (.32)	98 (.28)	78 (.27)	61 (.21)	20 (.07)	
C	70 (.24)	87 (.25)	50 (.17)	54 (.19)	23 (.08)	
D	63 (.21)	49 (.14)	23 (.08)	20 (.07)	14 (.05)	

Base N	A	B	C	D	N	P+
1502	695	352	284	169	32.24	.46
Item No.	N A	N B	N C	N D	P total	r ct
43	35.54	29.41	29.36	28.40	.99	.32

Choice	Low					High
	N 1	N 2	N 3	N 4	N 5	
A	65 (.22)	49 (.14)	40 (.14)	25 (.09)	10 (.04)	
*B	87 (.29)	141 (.41)	121 (.41)	164 (.56)	207 (.77)	
C	80 (.27)	92 (.27)	68 (.23)	49 (.17)	26 (.10)	
D	66 (.22)	63 (.18)	64 (.22)	53 (.18)	26 (.10)	

Base N	A	B	C	D	N	P+
1502	189	720	315	272	32.24	.48
Item No.	N A	N B	N C	N D	P total	r ct
44	28.70	34.91	29.77	30.51	.99	.26

	Choice	Low					High
		N 1	N 2	N 3	N 4	N 5	
	A	81 (.27)	72 (.21)	47 (.16)	34 (.12)	9 (.03)	
	*B	80 (.27)	189 (.55)	204 (.69)	229 (.79)	234 (.94)	
	C	73 (.25)	39 (.11)	26 (.09)	11 (.04)	6 (.02)	
	D	64 (.22)	47 (.14)	17 (.06)	16 (.06)	1 (.004)	
Base N 1502	A	243	956	155	145	32.24	P+ .64
Item No. 45	N A	28.27	N B 35.05	N C 26.63	N D 26.34	P total .99	r ct .42

	Choice	Low					High
		N 1	N 2	N 3	N 4	N 5	
	*A	69 (.23)	129 (.37)	125 (.43)	164 (.56)	165 (.61)	
	B	39 (.13)	27 (.08)	12 (.04)	7 (.02)	6 (.02)	
	C	85 (.29)	37 (.11)	38 (.13)	13 (.05)	2 (.01)	
	D	102 (.35)	155 (.45)	119 (.41)	108 (.37)	97 (.36)	
Base N 1502	A	652	B 91	C 175	D 581	N 32.27	P+ .43
Item No. 46	N A	34.88	N B 27.20	N C 26.42	N D 31.89	P total .99	r ct .22

	Choice	Low					High
		N 1	N 2	N 3	N 4	N 5	
	*A	161 (.54)	291 (.84)	259 (.88)	279 (.96)	266 (.99)	
	B	58 (.20)	35 (.10)	21 (.07)	7 (.02)	2 (.01)	
	C	29 (.10)	8 (.02)	4 (.01)	2 (.01)	0 (.000)	
	D	50 (.17)	14 (.04)	9 (.03)	4 (.01)	2 (.01)	
Base N 1502	A	1256	B 123	C 43	D 79	N 32.24	P+ .84
Item No. 47	N A	33.66	N B 25.83	N C 23.23	N D 24.59	P total .99	r ct .36

	Choice	Low					High
		N 1	N 2	N 3	N 4	N 5	
	A	24 (.08)	26 (.08)	12 (.04)	7 (.02)	1 (.004)	
	B	72 (.24)	46 (.13)	34 (.12)	19 (.07)	4 (.02)	
	C	83 (.28)	66 (.19)	42 (.14)	25 (.09)	4 (.02)	
	*D	119 (.40)	209 (.60)	206 (.70)	240 (.83)	261 (.97)	
Base N 1502	A	70	B 175	C 220	D 1035	N 32.24	P+ .69
Item No. 48	N A	26.99	N B 26.99	N C 27.46	N D 34.50	P total .99	r ct .37

Choice	Low					High	
	N 1	N 2	N 3	N 4	N 5		
A	58 (.20)	55 (.16)	31 (.11)	18 (.06)	11 (.04)		
B	52 (.18)	128 (.37)	160 (.54)	189 (.65)	221 (.82)		
C	65 (.22)	64 (.18)	34 (.12)	22 (.08)	10 (.04)		
D	122 (.41)	100 (.29)	69 (.24)	62 (.21)	28 (.10)		
Base N	150	750	195	381	32.25	P+	.50
Item No.	41	N B	N C	N D	P total	r st	
		35.74	27.91	29.34	.99	.38	

Choice	Low					High	
	N 1	N 2	N 3	N 4	N 5		
A	75 (.25)	53 (.15)	33 (.11)	22 (.08)	10 (.04)		
B	57 (.19)	50 (.14)	36 (.12)	19 (.07)	13 (.05)		
*C	99 (.33)	189 (.55)	187 (.64)	207 (.71)	220 (.82)		
D	67 (.23)	55 (.16)	38 (.13)	44 (.15)	27 (.10)		
Base N	1502	193	175	902	231	P+	.60
Item No.	50	N A	N B	N C	N D	P total	r st
		27.94	28.64	34.43	30.05	.99	.28

Choice	Low					High	
	N 1	N 2	N 3	N 4	N 5		
A	109 (.40)	143 (.44)	114 (.40)	98 (.34)	43 (.16)		
*B	75 (.28)	98 (.30)	99 (.34)	131 (.46)	182 (.68)		
C	34 (.13)	28 (.09)	20 (.07)	23 (.08)	19 (.07)		
D	55 (.20)	54 (.17)	55 (.19)	33 (.12)	22 (.08)		
Base N	150	585	124	219	32.47	P+	.41
Item No.	38	N B	N C	N D	P total	r st	
		35.38	30.54	30.19	.95	.23	

Choice	Low					High	
	N 1	N 2	N 3	N 4	N 5		
A	58 (.21)	30 (.09)	20 (.07)	20 (.07)	13 (.05)		
B	40 (.15)	30 (.09)	10 (.04)	13 (.05)	1 (.004)		
*C	105 (.38)	194 (.60)	237 (.82)	246 (.86)	251 (.94)		
D	71 (.26)	70 (.22)	22 (.08)	7 (.02)	2 (.01)		
Base N	1502	141	94	1033	172	P+	.71
Item No.	52	N A	N B	N C	N D	P total	r st
		28.31	26.81	34.65	25.81	.96	.39

Choice	Low					High
	N 1	N 2	N 3	N 4	N 5	
A	65 (.24)	76 (.23)	50 (.17)	36 (.13)	20 (.08)	
*B	45 (.16)	106 (.33)	140 (.49)	196 (.69)	213 (.80)	
C	88 (.32)	61 (.19)	39 (.14)	18 (.06)	12 (.01)	
D	76 (.28)	82 (.25)	59 (.21)	34 (.12)	20 (.08)	
Base N	A	B	C	D	M	P+
502	247	700	218	271	32.43	.49
Item No.	N A	N B	N C	N D	P total	ret
53	29.66	36.19	27.36	29.34	.96	.40

Choice	Low					High
	N 1	N 2	N 3	N 4	N 5	
*A	126 (.46)	198 (.61)	211 (.73)	233 (.84)	227 (.85)	
B	25 (.09)	8 (.03)	6 (.02)	3 (.01)	7 (.03)	
C	28 (.10)	23 (.07)	7 (.02)	4 (.01)	3 (.01)	
D	95 (.35)	95 (.29)	66 (.23)	40 (.14)	30 (.11)	
Base N	A	B	C	D	M	P+
1502	1001	49	65	326	32.46	.69
Item No.	N A	N B	N C	N D	P total	ret
54	34.08	27.80	26.77	29.33	.96	.24

Choice	Low					High
	N 1	N 2	N 3	N 4	N 5	
A	31 (.11)	15 (.05)	8 (.03)	5 (.02)	3 (.01)	
B	38 (.14)	30 (.09)	20 (.07)	4 (.01)	3 (.14)	
C	121 (.44)	120 (.37)	92 (.32)	96 (.34)	38 (.14)	
*D	83 (.30)	158 (.49)	169 (.59)	181 (.63)	223 (.84)	
Base N	A	B	C	D	M	P+
1502	62	95	467	814	32.47	.57
Item No.	N A	N B	N C	N D	P total	ret
55	26.40	27.14	30.12	34.90	.96	.28

Choice	Low					High
	N 1	N 2	N 3	N 4	N 5	
A	39 (.14)	43 (.13)	32 (.11)	21 (.07)	8 (.03)	
B	81 (.30)	72 (.22)	45 (.16)	27 (.10)	14 (.05)	
*C	82 (.30)	140 (.43)	160 (.56)	197 (.69)	227 (.85)	
D	71 (.26)	69 (.21)	50 (.17)	39 (.14)	17 (.06)	
Base N	A	B	C	D	M	P+
1502	143	239	806	246	32.45	.56
Item No.	N A	N B	N C	N D	P total	ret
56	29.24	28.33	35.21	29.29	.95	.32

Choice	Low					High
	N 1	N 2	N 3	N 4	N 5	
A	82 (.30)	37 (.11)	15 (.05)	4 (.01)	2 (.01)	
B	50 (.18)	55 (.17)	30 (.10)	37 (.13)	17 (.06)	
*C	101 (.37)	183 (.56)	188 (.65)	210 (.74)	215 (.81)	
D	40 (.15)	50 (.15)	57 (.20)	34 (.12)	33 (.12)	
Base N	A	B	C	D	N	P+
02	140	189	897	214	32.46	.62
Item No.	N A	N B	N C	N D	N total	rit
57	24.27	30.10	34.41	31.72	.96	.27

Choice	Low					High
	N 1	N 2	N 3	N 4	N 5	
A	71 (.26)	58 (.18)	49 (.17)	48 (.17)	51 (.19)	
*B	48 (.18)	45 (.14)	55 (.19)	51 (.18)	70 (.26)	
C	109 (.40)	171 (.53)	150 (.52)	156 (.55)	129 (.48)	
D	45 (.17)	51 (.16)	36 (.12)	31 (.11)	17 (.06)	
Base N	A	B	C	D	N	P+
1502	277	269	715	180	32.46	.19
Item No.	N A	N B	N C	N D	N total	rit
58	31.68	34.25	32.63	30.37	.96	.06

Choice	Low					High
	N 1	N 2	N 3	N 4	N 5	
A	35 (.13)	38 (.12)	24 (.08)	12 (.04)	6 (.02)	
B	27 (.10)	11 (.03)	13 (.05)	7 (.02)	0 (.000)	
C	55 (.20)	38 (.12)	30 (.10)	21 (.07)	6 (.02)	
*D	155 (.57)	238 (.73)	223 (.77)	246 (.86)	255 (.96)	
Base N	A	B	C	D	N	P+
502	115	58	150	1117	32.47	.77
Item No.	N A	N B	N C	N D	N total	rit
59	28.03	26.50	27.89	30.85	.96	.26

Choice	Low					High
	N 1	N 2	N 3	N 4	N 5	
A	59 (.22)	68 (.21)	72 (.25)	73 (.26)	89 (.33)	
B	55 (.20)	49 (.15)	37 (.13)	35 (.12)	13 (.05)	
C	61 (.22)	50 (.15)	30 (.10)	16 (.06)	14 (.05)	
*D	97 (.36)	158 (.49)	151 (.52)	162 (.57)	151 (.57)	
Base N	A	B	C	D	N	P+
1502	361	189	171	719	32.47	.50
Item No.	N A	N B	N C	N D	N total	rit
60	33.68	29.49	28.08	33.69	.96	.10

ภาคผนวก ก

แสดงกันดั้มของข้อคำถาม 59 ข้อ ในแบบทดสอบวัดผลสัมฤทธิ์วิชา

คณิตศาสตร์ ตามค่าความยากที่ประมาณภายใต้โมเดล 3PL

โมเดล 1PL และโมเดลดั้งเดิม

อันดับที่	โมเดล 3PL		โมเดล 1PL		โมเดลดั้งเดิม	
	ข้อที่	b	ข้อที่	b	ข้อที่	p
	17	-3.75	17	-2.68	17	.9401
	3	-2.92	3	-2.38	3	.9214
	9	-2.57	1	-2.29	1	.9154
	12	-2.00	9	-1.55	9	.8435
	1	-2.00	47	-1.54	47	.8389
	18	-1.99	7	-1.49	7	.8356
	7	-1.90	18	-1.37	18	.8182
	8	-1.63	12	-1.34	12	.8136
	47	-1.21	8	-1.20	8	.7943
0	59	-1.21	59	-1.07	59	.7725
1	19	-0.98	29	- .82	29	.7282
2	54	-0.77	19	- .81	19	.7257
3	29	-0.76	52	- .74	52	.7117
4	11	-0.65	11	- .66	11	.6984
5	52	-0.60	54	- .63	54	.6944
6	24	-0.50	48	- .63	48	.6911

อันดับที่	โมเดล 3PL		โมเดล 1PL		โมเดลดั้งเดิม	
	ข้อที่	b	ข้อที่	b	ข้อที่	p
17	32	-0.37	24	- .45	24	.6545
18	57	-0.28	31	- .38	31	.6418
19	45	-0.27	45	- .36	45	.6372
20	2	-0.23	32	- .30	32	.6245
21	48	-0.22	57	- .28	57	.6200
22	15	-0.15	2	- .25	2	.6145
23	50	-0.13	50	- .20	50	.6005
24	30	-0.04	23	- .18	23	.5972
25	31	-0.03	15	- .15	15	.5932
26	16	-0.02	16	- .11	30	.5836
27	40	-0.01	30	- .11	16	.5832
28	55	0.09	40	- .10	40	.5786
29	49	0.16	21	- .05	21	.5699
30	23	0.17	55	- .04	55	.5672
31	26	0.23	56	- .01	56	.5596
32	53	0.24	26	.11	26	.5313

อันดับที่	โมเดล 3PL		โมเดล 1PL		โมเดลดั้งเดิม	
	ข้อที่	b	ข้อที่	b	ข้อที่	p
33	56	0.31	20	.18	20	.5186
34	10	0.66	10	.19	10	.5140
35	43	0.78	60	.25	60	.5023
36	21	0.82	49	.26	49	.4993
37	20	0.85	53	.31	53	.4874
38	5	0.91	44	.36	44	.4800
39	44	0.93	43	.43	43	.4627
40	27	0.96	5	.54	5	.4348
41	46	1.07	46	.56	46	.4321
42	37	1.07	4	.60	4	.4254
43	4	1.13	27	.61	27	.4231
44	51	1.31	33	.68	33	.4075
45	60	1.34	51	.68	51	.4061
46	36	1.53	35	.69	35	.4048
47	33	1.57	36	.85	36	.3682
48	28	1.57	37	.94	37	.3509

อันดับที่	โมเดล 3PL		โมเดล 1PL		โมเดลดั้งเดิม	
	ข้อที่	b	ข้อที่	b	ข้อที่	p
49	38	1.60	28	1.11	34	.3149
50	42	1.68	34	1.12	28	.3138
51	39	1.80	38	1.23	42	.2929
52	13	2.08	42	1.23	38	.2923
53	6	2.18	25	1.35	25	.2723
54	22	2.18	13	1.35	13	.2710
55	34	2.62	22	1.44	22	.2543
56	35	2.83	6	1.62	6	.2244
57	25	2.87	39	1.73	39	.2071
58	58	2.91	41	1.86	41	.1884
59	41	3.91	58	1.90	58	.1853

ข้อ 14 มีค่า $p = .2437$