

นิพนธ์ต้นฉบับ

Original Article

การประยุกต์สมการถดถอยพหุพัวส์ของในงานวิจัย ทางด้านทันตสาธารณสุข

นิตยา พวงราช ส.ม. (ชีวสถิติ)

โรงพยาบาลคูเมือง อำเภอกูเมือง จังหวัดบุรีรัมย์

บทคัดย่อ การวิเคราะห์ข้อมูลที่มีลักษณะเป็นจำนวนนับโดยเฉพาะข้อมูลที่มีลักษณะการเกิดเหตุการณ์น้อย ๆ หรือข้อมูลที่มีค่าศูนย์จำนวนมาก หากใช้สถิติไม่ถูกต้องจะทำให้ค่าความคลาดเคลื่อนต่ำกว่าความเป็นจริง นำไปสู่การสรุปผลที่ผิดพลาด การศึกษานี้เป็นส่วนหนึ่งของการศึกษาเชิงวิเคราะห์แบบภาคตัดขวาง เรื่องความสัมพันธ์ระหว่างเงินค่าขนมกับปริมาณฟันผุ ถอน อุด ในเด็กอายุ 12 ปีที่ศึกษาในโรงเรียนประถมศึกษา อำเภอกูเมือง จังหวัดบุรีรัมย์ มีวัตถุประสงค์เพื่ออธิบายการประยุกต์สมการถดถอยพหุพัวส์ของในงานวิจัยทางด้านทันตสาธารณสุข กลุ่มตัวอย่างมี 299 คน สุ่มตัวอย่างด้วยวิธีการสุ่มแบบเป็นระบบ เก็บข้อมูลด้วยแบบสัมภาษณ์ แบบบันทึกเงินค่าขนม และแบบบันทึกการตรวจฟัน ผลการศึกษาพบ over-dispersion และมี zero outcome ร้อยละ 11.37 ทำการคัดเลือกสมการที่มีความเหมาะสมกับข้อมูลโดยนำสมการถดถอยพหุพัวส์แบบลบกรณีตัวแปรตามมีค่าศูนย์ เปรียบเทียบกับสมการถดถอยพหุพัวส์ของสกรณิตัวแปรตามมีค่าศูนย์ และนำสมการถดถอยพหุพัวส์แบบลบกรณีตัวแปรตามมีค่าศูนย์ เปรียบเทียบกับสมการถดถอยพหุพัวส์ของสกรณิตัวแปรตามมีค่าศูนย์ และดีกว่าสมการถดถอยพหุพัวส์แบบลบ ดังนั้น ควรประยุกต์สถิติขั้นสูงตัวอื่น เช่น Poisson data with correlation under Generalized Estimating Equation (GEE), Poisson data with nested data structures under Generalized Linear Mixed Model (GLMM), Generalized Additive Mixed Models (GAMM), Multilevel with multiple Poisson regression เพื่อพัฒนาคุณภาพงานวิจัยทางด้านทันตสาธารณสุขให้มีความก้าวหน้า

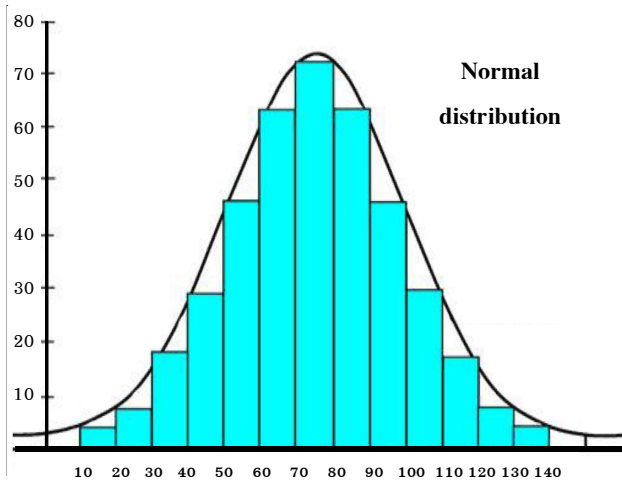
คำสำคัญ: พัวส์ของ, ทันตสาธารณสุข, วิจัย

บทนำ

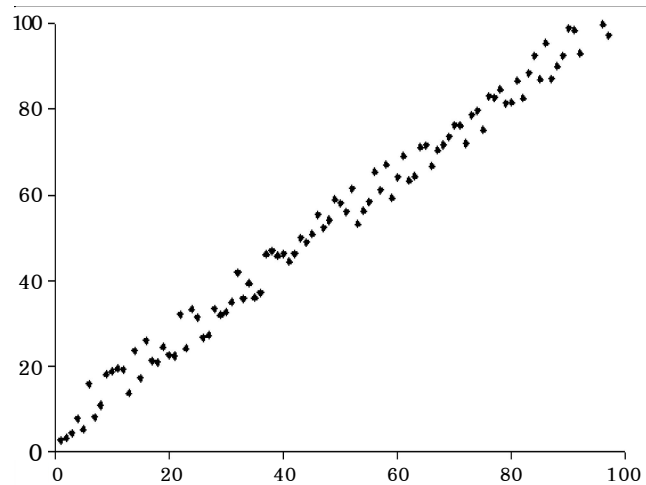
การวิเคราะห์ข้อมูลเพื่อหาปัจจัยที่มีความสัมพันธ์กับการเกิดฟันผุ พบว่า สถิติที่ใช้ในการวิเคราะห์ข้อมูลมีหลากหลายขึ้นอยู่กับตัวแปรผลลัพธ์ว่ามีลักษณะเป็นอย่างไร เช่น วัดฟันผุ กับ ไม่ผุ เป็นการเกิดเหตุการณ์เป็นแบบทวินาม (binary outcome) นักวิจัยจะวิเคราะห์ข้อมูลด้วยสถิติการถดถอยพหุโลจิสติกส์ (multiple logistic regression) หากนักวิจัยวัดค่าเฉลี่ยฟันผุ ถอน อุด (Decayed Missing Filled Teeth - DMFT) ของเด็กแต่ละคน

ตัวแปรผลลัพธ์มีลักษณะเป็นข้อมูลต่อเนื่อง (continuous outcome) นักวิจัยจะวิเคราะห์ข้อมูลด้วยสถิติการถดถอยพหุเชิงเส้น (multiple linear regression) อย่างไรก็ตาม การวิเคราะห์ข้อมูลด้วยสมการถดถอยพหุโลจิสติกส์ คือผลลัพธ์ที่ได้ไม่คำนึงถึงปริมาณฟันผุ ถอน อุด ของเด็กแต่ละคน ส่วนการวิเคราะห์ข้อมูลด้วยสมการถดถอยพหุเชิงเส้น มีเงื่อนไข คือ ค่าส่วนที่เหลือ (residual) ต้องมีการแจกแจงเป็นแบบปกติ (normality of residual) (ภาพที่ 1)⁽¹⁾ และลักษณะการกระจายของค่าส่วนที่เหลือ (re-

ภาพที่ 1 เงื่อนไขการวิเคราะห์ข้อมูลด้วยสถิติการถดถอย-
พหุเชิงเส้น: normality of residual⁽¹⁾



ภาพที่ 2 เงื่อนไขการวิเคราะห์ข้อมูลด้วยสถิติการถดถอย-
พหุเชิงเส้น: homoscedasticity⁽²⁾



sidual) ในทุกๆ ระดับของค่าคาดหวัง ($\hat{y}_i$) ต้องคงที่ (homoscedasticity) (ภาพที่ 2)⁽²⁾ อย่างไรก็ตาม เงื่อนไขดังกล่าวอาจไม่เหมาะสมกับข้อมูลที่มีลักษณะเป็นจำนวนนับ (count data) โดยเฉพาะข้อมูลที่มีการเกิดเหตุการณ์น้อยๆ (rare event) หรือ ข้อมูลที่มีค่าศูนย์จำนวนมาก (zero outcome) เพราะทำให้ค่าความคลาดเคลื่อนต่ำกว่าความเป็นจริง นำไปสู่การสรุปผลที่ผิดพลาดได้^(3,4) และทำให้ตัวประมาณไม่มีประสิทธิภาพ ไม่คงเส้นคงวา มีความเอนเอียง⁽⁵⁾ การศึกษานี้มีวัตถุประสงค์เพื่ออธิบายการประยุกต์สมการถดถอยพหุปัจจัยในงานวิจัยทางด้านทันตสาธารณสุข เพื่อพัฒนางานวิจัยให้มีคุณภาพยิ่งขึ้นต่อไป

วิธีการศึกษา

ผลงานนี้เป็นส่วนหนึ่งของการศึกษาเรื่อง ความสัมพันธ์ระหว่างเงินค่าขนมกับปริมาณฟันผุ ถอน อุด ในเด็กอายุ 12 ปี ที่ศึกษาในโรงเรียนประถมศึกษา อำเภอคูเมือง จังหวัดบุรีรัมย์ ซึ่งผ่านการรับรองจากคณะกรรมการจริยธรรมการวิจัยในมนุษย์จากมหาวิทยาลัยขอนแก่น เป็นการศึกษาเชิงวิเคราะห์แบบภาคตัดขวาง (analytical cross sectional study) ผู้วิจัยวิเคราะห์ข้อมูลเพื่อตอบคำถามตามวัตถุประสงค์การวิจัย จึงใช้สมการถดถอยพหุปัจจัย ในการประมาณขนาดตัวอย่างสำหรับการวิเคราะห์ด้วย

วิธีการของ Signorini⁽⁶⁾ ปรับขนาดตัวอย่างตามการวิเคราะห์ถดถอยพหุปัจจัย⁽⁷⁾ และค่า item non response rate อีกครั้ง ดังนั้น กลุ่มตัวอย่างที่ศึกษาในครั้งนี้เท่ากับ 299 คน ใช้วิธีการสุ่มตัวอย่างแบบมีระบบ (systematic random sampling) เก็บรวบรวมข้อมูลโดยใช้แบบสัมภาษณ์ แบบบันทึกเงินค่าขนม แบบบันทึกการตรวจฟัน ปรับมาตรฐานการตรวจฟันกับทันตแพทย์จำแนกโมเดลตามตัวแปรอิสระที่สนใจศึกษาได้แก่ โมเดลรายรับ (model 1) โมเดลรายจ่ายค่าอาหารประเภทแป้งและน้ำตาลทุกชนิด (model 2) และโมเดลรายจ่ายเฉพาะค่าขนม (model 3)

เป้าหมายของการวิเคราะห์ถดถอยพหุปัจจัย คือ หาความสัมพันธ์ระหว่างตัวแปรตามที่มีลักษณะของข้อมูลเป็นจำนวนนับ (count data) กับตัวแปรอิสระในโมเดลตั้งแต่ 2 ตัวแปรขึ้นไป โดยควบคุมอิทธิพลของปัจจัยอื่นไปพร้อมๆ กัน ทั้งนี้ ตัวแปรอิสระสามารถเป็นข้อมูลเชิงคุณภาพหรือข้อมูลเชิงปริมาณก็ได้ อย่างไรก็ตาม สมการถดถอยพหุปัจจัยของสมิฟังก์ชันการแจกแจงหลายรูปแบบ งานวิจัยนี้ขอเสนอเพียง 4 รูปแบบ คือ สมการถดถอยพหุปัจจัย (multiple Poisson regression) ซึ่งมีเงื่อนไขคือค่าความแปรปรวนของตัวแปรตามต้องเท่ากับค่าเฉลี่ยเรียกว่า equi-dispersion อย่างไรก็ตาม ข้อมูลพื้นฐานส่วนใหญ่มักพบว่าความแปรปรวนมีค่า

สูงกว่าค่าเฉลี่ย เรียกว่า over-dispersion นักวิจัยสามารถเลือกใช้สมการถดถอยพหุพัวส์ซองแบบลบ (negative binomial regression) มาเป็นเครื่องมือในการแสวงหาคำตอบ สมการถดถอยพหุพัวส์ซองแบบลบที่นิยมใช้มีอยู่ 2 รูปแบบ คือ สมการที่มีความแปรปรวนเป็นฟังก์ชันเชิงเส้นตรงกับค่าเฉลี่ย (linear mean variance negative binomial) เรียกว่า NB1 และ สมการพหุพัวส์ซองแบบลบที่มีความแปรปรวนเป็นฟังก์ชันกำลังสองของค่าเฉลี่ย (quadratic mean variance negative binomial) เรียกว่า NB2 การวิเคราะห์ข้อมูลด้วยสมการถดถอยพหุพัวส์ซอง (multiple Poisson regression) สามารถตรวจสอบการเกิด over-dispersion ระหว่างสมการถดถอยพหุพัวส์ซอง กับสมการถดถอยพหุพัวส์ซองแบบลบด้วยสถิติหลายตัวได้แก่ score test⁽⁸⁾ และ likelihood ratio test⁽⁵⁾ นอกจากนี้ข้อสรุปที่ได้อาจไม่ถูกต้องหากตัวแปรตามมีค่าเป็นศูนย์ สมการถดถอยพหุพัวส์ซองกรณีตัวแปรตามมีค่าศูนย์เป็นอีกหนึ่งทางเลือกที่สามารถนำมาใช้ในการแก้ไขปัญหากรณีตัวแปรตามเกิดปัญหาความแปรปรวนมีค่ามากกว่าค่าเฉลี่ยร่วมกับตัวแปรตามมีค่าศูนย์ หากนักวิจัยใช้สมการถดถอยพหุพัวส์ซองแบบลบหรือสมการถดถอยพหุพัวส์ซองกรณีตัวแปรตามมีค่าศูนย์ จะทำให้การประมาณค่าพารามิเตอร์ของตัวแปรอิสระที่มีอยู่ในโมเดล

ผิดพลาด นักวิจัยสามารถใช้สมการถดถอยพหุพัวส์ซองแบบลบกรณีตัวแปรตามมีค่าศูนย์มาเป็นเครื่องมือในการแสวงหาคำตอบได้ รายละเอียดนำเสนอในตารางที่ 1

ผลการศึกษา

การศึกษาความสัมพันธ์ระหว่างเงินค่าขนมกับปริมาณฟันผุ ถอน ในเด็กอายุ 12 ปี ที่ศึกษาในโรงเรียนประถมศึกษา อำเภอคูเมือง จังหวัดบุรีรัมย์ ผู้วิจัยตรวจสอบการเกิด over-dispersion ระหว่างสมการถดถอยพหุพัวส์ซอง กับ สมการถดถอยพหุพัวส์ซองแบบลบ พบว่าเกิด over-dispersion ในโมเดลที่ 1, 2 และ 3 รายละเอียดนำเสนอในตารางที่ 2

เมื่อเกิดปัญหา over-dispersion ในโมเดล 1, 2 และ 3 และกลุ่มตัวอย่างเด็กอายุ 12 ปี ที่ศึกษาในโรงเรียนประถมศึกษา อำเภอคูเมือง จังหวัดบุรีรัมย์ ปราศจากโรคฟันผุ (caries free) เพียงร้อยละ 11.37 จึงประเมินความเหมาะสมของโมเดลเบื้องต้นด้วยค่า Akaike Information Criterion (AIC) แล้วเปรียบเทียบความเหมาะสมของสมการที่นำมาใช้วิเคราะห์ข้อมูลเมื่อตัวแปรตามมีค่าศูนย์และเกิดปัญหา over-dispersion ระหว่างสมการถดถอยพหุพัวส์ซองแบบลบกรณีตัวแปรตามมีค่าศูนย์กับสมการถดถอยพหุพัวส์ซองกรณีตัวแปรตามมีค่าศูนย์

ตารางที่ 1 สมการถดถอยพหุพัวส์ซองที่ใช้ในการวิเคราะห์ข้อมูลการเกิดโรคฟันผุ

สถิติ	การแจกแจง	รูปแบบโมเดล	over dispersion	ตัวแปรตามมีค่าศูนย์
Poisson regression	Poisson	Poisson	—	—
Negative binomial regression				
– NB1	Poisson + gamma	negative binomial	++	—
– NB2				
Zero inflated Poisson regression (ZIP)	Poisson + binomial	– Poisson + logistic – Poisson + probit	—	++
Zero inflated negative binomial regression (ZINB)	gamma + binomial	– negative binomial + logistic – probit + logistic	++	++

ถดถอยพหุวินามแบบลบในทุกๆ โมเดล การศึกษา
เลือกสมการถดถอยพหุวินามแบบลบกรณีตัวแปรตาม
มีค่าศูนย์ มาใช้ในการวิเคราะห์ข้อมูล พร้อมทั้งนำเสนอ
ตัวอย่างรายงานวิจัยที่ใช้สถิติการถดถอยพหุวิสรของ
วิเคราะห์ปัจจัยที่มีความสัมพันธ์กับปริมาณพันธุ์ ถอน อุด
ในพื้นน้ำนมและพันธุกรรมรายละเอียดในตารางที่ 3, 4
และ 5

ตรา

โมเดล	สถิติ score				สถิติ likelihood ratio	
	NB1	p-value	NB2	p-value	G ²	p-value
1	168.21	<0.0001	163.42	<0.0001	102.26	<0.0001
2	157.72	<0.0001	147.06	<0.0001	96.79	<0.0001
3	139.34	<0.0001	135.98	<0.0001	88.78	<0.0001

ตรา

โมเดล	ค่า AIC			สถิติ likelihood ratio		สถิติ vuong	
				ZINB : ZIP		ZINB : NB	
	ZIP	NB	ZINB	G ²	p-value	Z	p-value
1	4.93	4.85	4.81	37.82	<0.0001	1.81	0.035
2	4.91	4.84	4.80	32.40	<0.0001	1.97	0.024
3	4.89	4.83	4.79	33.32	<0.0001	1.84	0.033
ความเหมาะสม				ZINB เหมาะสมกว่า ZIP		ZINB เหมาะสมกว่า NB	

ตารา

โมเดล	สถิติ NB			สถิติ ZINB		
	ϕ	95%CI	p-value	ϕ	95% CI	p-value
1	0.29	0.21 – 0.41	<0.0001	0.16	0.10 – 0.26	<0.0001
2	0.28	0.20 – 0.39	<0.0001	0.14	0.08 – 0.24	<0.0001
3	0.27	0.19 – 0.38	<0.0001	0.15	0.09 – 0.25	<0.0001

วิจารณ์

เพิ่มการนำสมการดังกล่าวไปประยุกต์ในโรคอื่น ๆ ที่มีลักษณะการเกิดโรคที่คล้ายคลึงกัน เช่น กลุ่มโรคที่เกิดจากหลายปัจจัย (multifactorial disease) และสอดคล้องกับผลการศึกษาของ Böhning และคณะ⁽¹⁰⁾ ที่เสนอแนะว่าการประยุกต์สมการถดถอยพหุปัจจัยของกรณีตัวแปรตามมีค่าศูนย์ สามารถนำมาใช้อธิบายความสัมพันธ์ระหว่างตัวแปรอิสระต่างๆ ที่เกี่ยวข้องกับงานส่งเสริมทันตสุขภาพหรือประเมินผล โครงการในงานป้องกันโรคต่างๆ ในพื้นที่ได้เป็นอย่างดี

สรุป

การประยุกต์สมการถดถอยพหุพหุนามแบบลบกรณีตัวแปรตามมีค่าศูนย์มาใช้เป็นเครื่องมือในการแสวงหาคำตอบพบน้อยมาก โรคพันธุเกิดจากหลายปัจจัย (Multifactorial Disease) และระบาดวิทยาของการเกิดโรคพันธุมีความสัมพันธ์กับปัจจัยส่วนบุคคล ครอบครัว ชุมชน และสิ่งแวดล้อม ดังนั้น การศึกษาระบาดของโรคพันธุควรมีการประยุกต์สถิติขั้นสูงตัวอื่น เช่น Poisson data with correlation under Generalized Estimating Equation (GEE), Poisson data with nested data structures under Generalized Linear Mixed Model (GLMM), Generalized Additive Mixed Models (GAMM), Multilevel with multiple Poisson regression เพื่อพัฒนาคุณภาพงานวิจัยทางด้านทันตสาธารณสุขและวิทยาศาสตร์สุขภาพสาขาอื่นให้มีความก้าวหน้ายิ่งขึ้นต่อไป

กิตติกรรมประกาศ

ขอขอบคุณสำนักงานกองทุนสนับสนุนการสร้างเสริมสุขภาพ สำนักงานพัฒนานโยบายสุขภาพระหว่างประเทศ และแผนยุทธศาสตร์การพัฒนาศักยภาพประชากรและสุขภาพของประชากรไทย ที่สนับสนุนทุนในการทำวิจัย และขอขอบคุณผู้ที่เกี่ยวข้องทุกท่าน ที่ให้ความร่วมมือและสนับสนุนการดำเนินงานวิจัยนี้จนสำเร็จลุล่วงด้วยดี

เอกสารอ้างอิง

- สรุป**

การประยุกต์สมการถดถอยพหุพหุนามแบบลบกรณีตัวแปรตามมีค่าศูนย์มาใช้เป็นเครื่องมือในการแสวงหาคำตอบพบน้อยมาก โรคฟันผุเกิดจากหลายปัจจัย (Multifactorial Disease) และระบาดวิทยาของการเกิดโรคฟันผุมีความสัมพันธ์กับปัจจัยส่วนบุคคล ครอบครัว ชุมชน และสิ่งแวดล้อม ดังนั้น การศึกษาระบาดของโรคฟันผุควรมีการประยุกต์สถิติขั้นสูงตัวอื่น เช่น Poisson data with correlation under Generalized Estimating Equation (GEE), Poisson data with nested data structures under Generalized Linear Mixed Model (GLMM), Generalized Additive Mixed Models (GAMM), Multilevel with multiple Poisson regression เพื่อพัฒนาคุณภาพงานวิจัยทางด้านทันตสาธารณสุขและวิทยาศาสตร์สุขภาพสาขาอื่นให้มีความก้าวหน้ายิ่งขึ้นต่อไป

กิตติกรรมประกาศ

ขอขอบคุณสำนักงานกองทุนสนับสนุนการสร้างเสริมสุขภาพ สำนักงานพัฒนานโยบายสุขภาพระหว่างประเทศ และแผนยุทธศาสตร์การพัฒนาดัชนีประเมินภาระโรคและสุขภาพของประชากรไทย ที่สนับสนุนทุนในการทำวิจัย และขอขอบคุณผู้ที่เกี่ยวข้องทุกท่านที่ให้ความร่วมมือและสนับสนุนการดำเนินงานวิจัยนี้จนสำเร็จลุล่วงด้วยดี

เอกสารอ้างอิง

 - MathBitsNotebook.com. Normal distribution [Internet]. 2016 [cited 2016 Nov 9]. Available from: <http://mathbitsnotebook.com/Algebra2/Statistics/STnormalDistribution.html>
 - Wikipedia. Homoscedasticity [Internet]. 2016 [cited 2016 Nov 9]. Available from: <https://en.wikipedia.org/wiki/Homoscedasticity#/media/File:Homoscedasticity.png>
 - Gardner W, Mulvey EP, Shaw EC. Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. Psychol Bull 1995;118:392–404.
 - Cohen J, Cohen P, West S, Aiken L. Applied multiple regression/correlation analysis for the behavioral sciences. 3rd ed. Mahwah, New Jersey: Lawrence Erlbaum Associates Pub; 2003.
 - Long JS. Regression models for categorical and limited dependent variable. Thousand Oaks, California: SAGE Pub; 1997.
 - Signorini DF. Sample size for Poisson regression. Biometrika 1991;78:446–50.
 - Hsieh YF, Block AD, Larsen DM. A simple method of sample size calculation for linear and logistic regression. Stat Med 1998;17:1623–34.
 - Dean CB. Testing for over-dispersion in Poisson and binomial regression models. American Journal of Statistics Association 1992;87:451–7.
 - Lewsey JD, Thomson WM. The utility of the zero-inflated Poisson and zero-inflated negative binomial models: A case study of cross-sectional and longitudinal DMF data examining the effect of socio-economic status. Community Dent Oral Epidemiol 2004;32:183–9.
 - Böhning D, Dietz E, Schlattmann P, Mendonça L, Kirchner U. The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. Journal of the Royal Statistical Society 1999;162:195–209.
 - Tramini P, Molinari N, Tentscher M, Demattei C, Schulte AG. Association between caries experience and body mass index in 12-year-old French children. Caries Res 2009;43:468–73.
 - Solinas G, Campus G, Maida C, Sotgiu G, Cagetti MG, Lesaffre E, Castiglia P. What statistical method should be used to evaluate risk factors associated with dmfs index? Evidence from the National Pathfinder Survey of 4-year-old Italian children. Community Dent Oral Epidemiol 2009;37:539–46.

Abstract: Applied Multiple Poisson Regression for Dental Epidemiology Research

Nittaya Phuangrach, M.P.H. (Biostatistics)

Khumuang Hospital, Khumuang District, Buriram Province

Journal of Health Science 2016;25:1033-9.

In considering data as count data analysis especially for rare event and zero outcome, if the researcher analyzed data with inappropriate statistics, it will be very likely to get wrong evidence and errors leading to wrong conclusions. This study is a part of the analytical cross sectional study on “the association between daily pocket money and decayed, missing and filled teeth among children aged 12 Years in primary school at Khumuang District, Buriram Province”. The purpose of this study was to explore the application of multiple Poisson regression for dental epidemiology research. Samples were 299 students recruited by a systematic sampling method. An oral health survey form, and daily pocket money compilation forms and questionnaires for face-to-face interview were employed for data collection. It was found that there was 11.37% of over-dispersion problem and zero outcomes. Researcher selected the best statistics for analyzed the data by comparable mixed two processes between zero inflated negative binomial regression and zero inflated Poisson regression, and comparable mixed distributions between zero inflated negative binomial regression and negative binomial regression the results were found zero inflated negative binomial regression was appropriated more than zero inflated Poisson regression and better than negative binomial regression. There for advance statistics should be applied, for example correlation under Generalized Estimating Equation (GEE), Poisson data with nested data structures under generalized linear mixed model (GLMM), generalized additive mixed models (GAMM), multilevel with multiple Poisson regression to developed quality of dental epidemiology and health science research.

Key words: Poisson, dental epidemiology, research