หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

บทนำ

1.1 ความสำคัญและความเป็นมาของปัญหา

การวิจัยสาขาต่าง ๆ ทางด้านสังคมศาสตร์และวิทยาศาสตร์จำเป็นต้องอาศัยวิธีการทางด้านสถิติเพื่อหาผลสรุปของสมมุติฐานที่ผู้วิจัยตั้งขึ้น ปัญหาข้อหนึ่งซึ่งผู้วิจัยส่วนใหญ่ประสบกันมากคือ การเลือกใช้ตัวสถิติทดสอบที่เหมาะสมกับข้อมูลที่นำมาวิเคราะห์เพราะว่าข้อมูลที่นำมาวิเคราะห์ต้องสอดคล้องกับวิธีการทางสถิติและข้อตกลงเบื้องต้นของตัวสถิติที่ใช้ในการวิเคราะห์ข้อมูลนั้นซึ่งแบ่งเป็น 2 ประเภทคือ

ก. <u>สถิติที่ใช้พารามิเตอร์</u>(Parametric Statistics) ผู้วิจัยจะใช้เพื่อวิเคราะห์ข้อมูล เมื่อตัวอย่างที่สุ่มมาจากประชากรซึ่งทราบว่ามีการแจกแจงอย่างใดอย่างหนึ่ง ผู้วิจัยจะประมาณค่าของพารามิเตอร์ที่ไม่ทราบจากค่าสถิติ (Statistics) ซึ่งคำนวนจากตัวอย่างโดยสุ่มจากประชากร

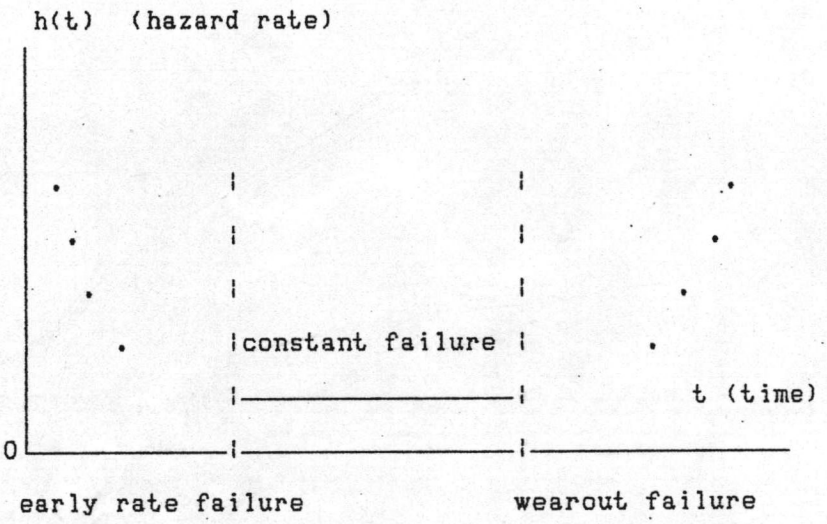
ข. <u>สถิติที่ไม่ใช้พารามิเตอร์</u>(Nonparametric Statistics)ผู้วิจัยจะใช้วิธีการนี้เมื่อไม่สามารถระบุได้ว่าข้อมูลที่นำมาวิเคราะห์มีการแจกแจงเป็นแบบใดวิธีการนี้ใช้ได้ง่ายและสะดวกเพราะไม่ต้องคำนึงถึงข้อตกลงเบื้องต้นของประชากรมากนัก จึงเป็นที่นิยมโดยทั่วไป

ในกรณีที่ผู้วิจัยสนใจศึกษาการวิเคราะห์ข้อมูลเชิงปริมาณ (quantitative data) บางประเภทเช่น ข้อมูลที่สนใจเป็นเวลาที่ต้องรอในเรื่องความน่าจะเป็น ข้อมูลเป็นเวลาที่เข้ามารับบริการ(inter-arrival time) ในระบบแถวคอย ข้อมูลเป็นข้อมูลอายุ (life data) เช่น อัตราการเสีย (hazard rate หรือ hazard function)ในช่วงเวลาหนึ่งของเครื่องจักรในงานวิศวกรรม หรือ อัตราการตาย(mortality rate)ในเรื่องความเที่ยง(reliability) ฯลฯ

ในทางทฤษฎีสามารถพิสูจน์ได้ว่า ข้อมูลเหล่านี้มีการแจกแจงแบบเอกซ์โปเนนเชียล (exponential distribution) ในทางปฏิบัตินักวิจัยที่จะนำข้อมูลเหล่านี้มาทำการวิเคราะห์ทางสถิติควรที่จะได้มีการทดสอบการแจกแจงของข้อมูลเสียก่อน เพราะถ้าคุณสมบัติของข้อมูลที่นำมาวิเคราะห์มีการแจกแจงที่ไม่เป็นไปตามข้อกำหนด ความถูกต้องและความหมายของผลที่ได้จากการวิเคราะห์ด้วยวิธีการทางสถิติประเภทจำกัดรูปแบบการแจกแจงซึ่งขึ้นอยู่กับความถูกต้องของข้อกำหนดในตอนแรก ก็จะทำให้ผลที่ได้จากการวิเคราะห์ไม่ถูกต้อง ในการศึกษาว่าข้อมูลมี

การแจกแจงแบบเอกซ์โปเนนเชียลหรือไม่จะทำได้โดยการเปรียบเทียบการแจกแจงของข้อมูลที่วิเคราะห์กับการแจกแจงของข้อมูลที่คิดว่าควรจะเป็น(theoretical distribution)แล้วใช้ตัวสถิติซึ่งนักสถิติหลายท่านได้คิดขึ้นทำการทดสอบ

Epstein(ค.ศ.1960)ได้คิดตัวสถิติที่ใช้ทดสอบอัตราเสียที่คงที่(constant failure rate) ซึ่งพบมากในเรื่องข้อมูลอายุ (life data) โดยการแจกแจงที่ Epstein สนใจคือการแจกแจงแบบเอกซ์โปเนนเชียล ซึ่งรูปแบบอัตราเสียคงที่สามารถพิจารณาได้จากรูปต่อไปนี้



รูปที่ 1.1.1 แสดงอัตราการเสียของข้อมูลอายุ

$$h(t) = f(t)/(1-F(t))$$

เมื่อ $F(t)$ = ฟังก์ชันการแจกแจงสะสม

$f(t)$ = ฟังก์ชันความหนาแน่น

กรณีที่ $f(t)$ = ฟังก์ชันความหนาแน่นแบบเอกซ์โปเนนเชียล

จะได้ว่า $f(t) = (1/\beta)*\exp(-t/\beta)$, $t > 0$

$F(t) = 1-\exp(-t/\beta)$, $t > 0$

ดังนั้น $h(t) = [(1/\beta)*\exp(-t/\beta)] / [1-(1-\exp(-t/\beta))]$

$= [(1/\beta)*\exp(-t/\beta)] / \exp(-t/\beta)$

$= 1/\beta$ (ค่าคงที่)

กล่าวคือ อัตราการเสียของการแจกแจงแบบเอกซ์โปเนนเชียลมีค่าคงที่

ต่อมานักสถิติหลายท่านได้คิดตัวสถิติอื่นๆ ที่ใช้ทดสอบการแจกแจงแบบเอกซ์โปเนนเชียล ซึ่งแบ่งได้เป็น 3 ประเภทคือ

ก. ทดสอบโดยใช้ Regression of the order statistics on some parameter of the distribution of the order statistics ได้แก่ Regresstion Test, Shapiro and Wilk Test และ Probability Plot Correlation Coefficient

ข. ทดสอบโดยใช้ empirical distribution function ได้แก่ Kolmogorov-Smirnov Test, Cramer-Von-Mises Test ฯลฯ

ค. ทดสอบโดยใช้ Weighted difference of sucessive order statistics ได้แก่ Gnedenko F Test และ Bivariate F Test

1.2 วัตถุประสงค์ของการวิจัย

ในการวิจัยครั้งนี้ต้องการเปรียบเทียบอำนาจการทดสอบของตัวสถิติที่ใช้ทดสอบการแจกแจงแบบเอกซ์โปเนนเชียล 3 ตัวคือ

ก. Regression Test (Z)

ข. Kolmogorov-Smirnov Test (K)

ค. Gnedenko F Test (F)

โดยทำการทดสอบภายใต้การวิเคราะห์ข้อมูลทั้งหมด (complete data) และการวิเคราะห์ข้อมูลที่มีข้อมูลขาดหาย (censored data)

1.3 สมมุติฐานของการวิจัย

ก. ลักษณะการแจกแจงของประชากรและขนาดตัวอย่างเดียวกันจะมีผลทำให้ตัวสถิติทดสอบ Z มีอำนาจการทดสอบสูงสุด

ข. ขนาดตัวอย่างและเปอร์เซนต์ข้อมูลขาดหายจะมีผลทำให้อำนาจการทดสอบของตัวสถิติทดสอบ Z สูงสุด

1.4 ข้อตกลงเบื้องต้น

ในการวิจัยครั้งนี้จะถือว่า อำนาจการทดสอบ (power of test) และความสามารถในการควบคุมความคลาดเคลื่อนประเภทที่ 1 เป็นดัชนีสำคัญที่จะใช้เป็นเกณฑ์ในการเลือกตัวสถิติทดสอบ

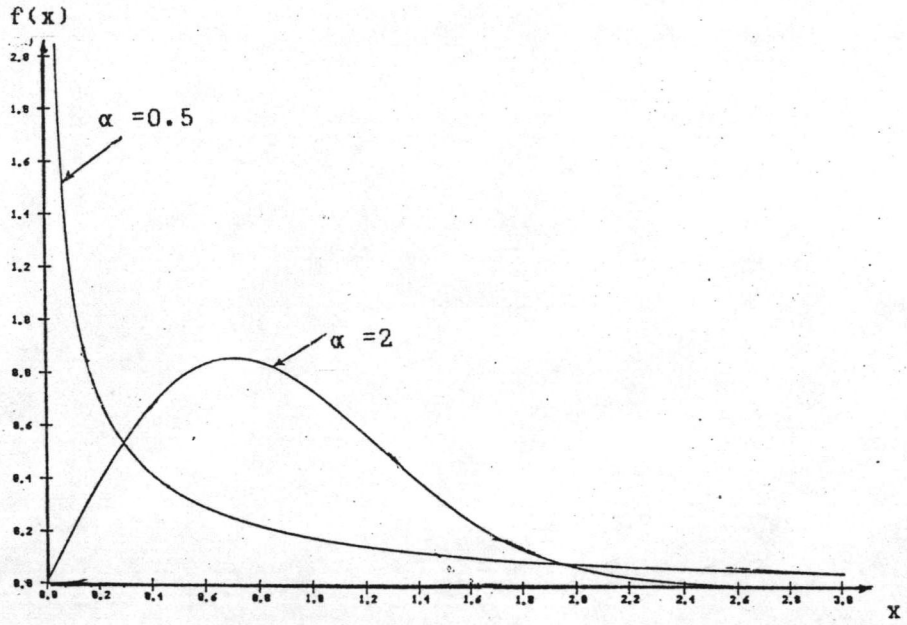
1.5 ขอบเขตการวิจัย

1.5.1 ศึกษาอำนาจการทดสอบและความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ของตัวสถิติ 3 ตัว โดยที่ประชากรมีการแจกแจงดังนี้

ก. การแจกแจงแบบไวบูลล์ (Weibull distribution)
ฟังก์ชันความหนาแน่นอยู่ในรูปของ

$$f(x) = \begin{cases} \dfrac{\alpha\, x^{\alpha-1}\, .\exp[-(x/\beta)^{\alpha}]}{\beta^{\alpha}} & , 0 < x < \infty\ ,\alpha > 0,\ \beta < 0 \\ 0 & , \text{อื่น ๆ} \end{cases}$$

ผู้วิจัยสนใจศึกษาเมื่อ scale parameter(β) เท่ากับ 1 และ shape parameter (α) เท่ากับ 0.5 และ 2.0 ตามลำดับ ซึ่งสามารถเขียนรูปได้ดังนี้



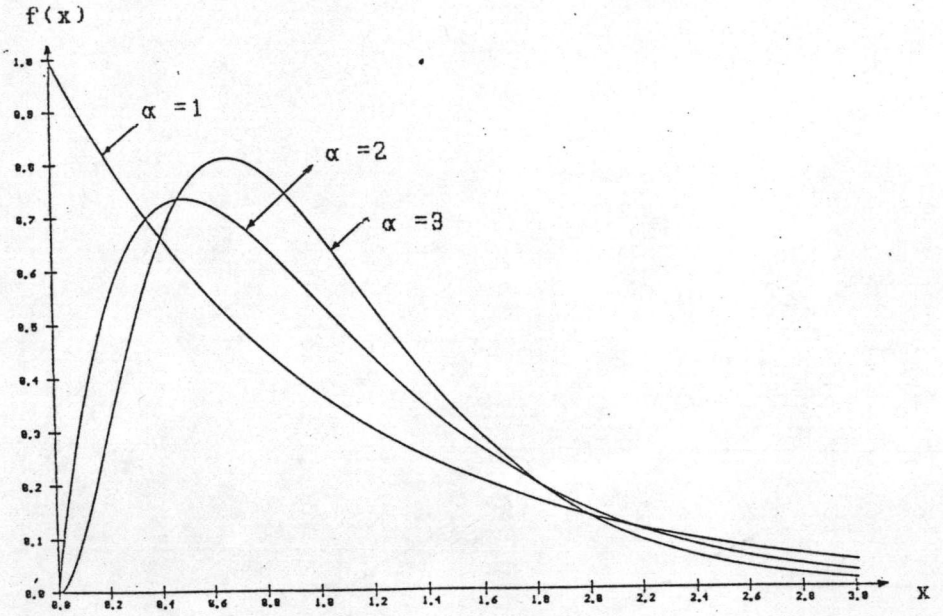
รูปที่ 1.5.1.1
แสดงการแจกแจงแบบไวบูลล์ เมื่อ $\beta = 1, \alpha = 0.5$ และ 2.0 ตามลำดับ

ข. การแจกแจงแบบแกมมา (Gamma distribution)
ฟังก์ชันความหนาแน่นอยู่ในรูปของ

$$f(x) = \begin{cases} \dfrac{x^{\alpha-1}.\exp(-x/\beta)}{\beta^{\alpha}.\Gamma(\alpha)} & , x > 0 , \alpha > 0 , \beta > 0 \\ 0 & , \text{อื่น ๆ} \end{cases}$$

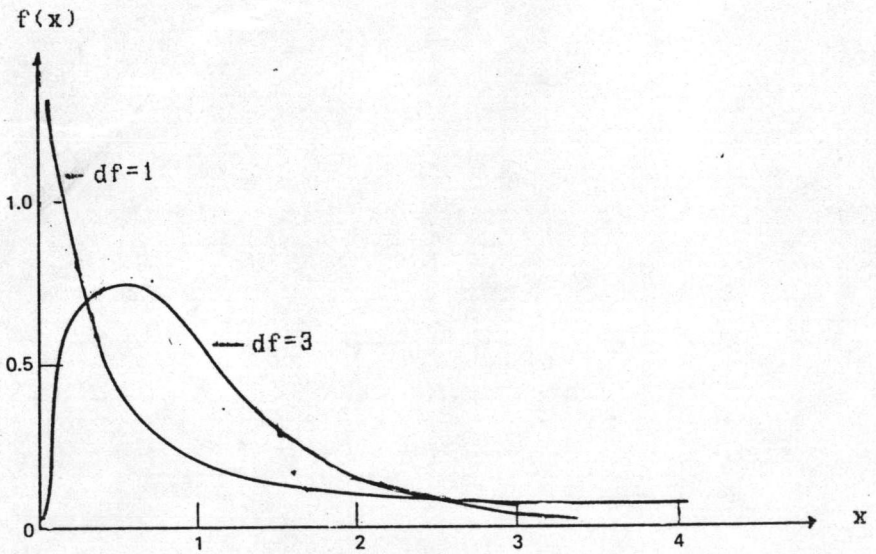
ผู้วิจัยสนใจศึกษาเมื่อ scale parameter(β) =1 และ shape parameter(α) =1,2 และ 3 ตามลำดับ ซึ่งในกรณีที่ $\alpha = 1$ การแจกแจงแบบแกมมาคือการแจกแจงแบบเอกซ์โปเนนเชียล และเมื่อ $\beta = 2, \alpha = n/2$ การแจกแจงแบบแกมมาคือ การแจกแจงแบบไคสแควร์ ด้วยระดับความเป็นอิสระ n ซึ่งสามารถเขียนรูปได้ดังนี้

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 1.5.1.2

แสดงการแจกแจงแบบแกมมา เมื่อ β = 1 α = 1,2 และ 3 ตามลำดับ



รูปที่ 1.5.1.3

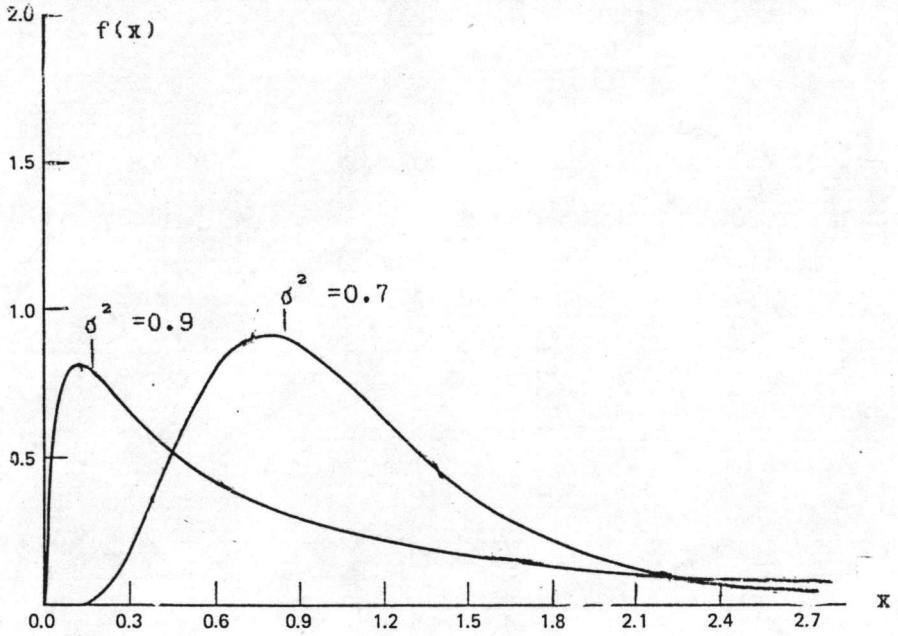
แสดงการแจกแจงแบบการแจกแจงแบบไคสแควร์ เมื่อระดับความเป็นอิสระ 1 และ 3 ตามลำดับ

ค. <u>การแจกแจงแบบลอกนอร์มอล</u> (Lognormal distribution)

ฟังก์ชันความหนาแน่นอยู่ในรูปของ

$$f(x) = \begin{cases} \frac{1}{x.\sigma.\sqrt{2\pi}} \exp\left[-(\ln x - \mu)^2 / 2\sigma^2\right] & , x > 0, \sigma > 0 \\ & , -\infty < \beta < \infty \\ 0 & , \text{อื่น ๆ} \end{cases}$$

ผู้วิจัยสนใจศึกษา ณ ค่าเฉลี่ย(μ)เท่ากับ 0 และ ความแปรปรวน(σ^2)เท่ากับ 0.7 และ0.9 ตามลำดับ สำหรับรูปแบบการแจกแจงแบบลอกนอร์มอลสามารถเขียนรูปได้ดังนี้



<u>รูปที่ 1.5.1.4</u>

แสดงการแจกแจงแบบลอกนอร์มอล เมื่อพารามิเตอร์ μ = 0 และ σ^2 = 0.7 และ 0.9 ตามลำดับ

จากรูปแบบการแจกแจงทั้งหมดจะเห็นว่ามีการแจกแจงแบบเบ้ขวา ซึ่งพบมากในทางปฏิบัติ เช่น ข้อมูลที่เกี่ยวกับอายุ และข้อมูลทางด้านแพทย์

1.5.2 ตัวสถิติที่ใช้ทดสอบการแจกแจงแบบเอกซ์โปเนนเชียล กรณีมีข้อมูลขาดหายจะใช้ตัวสถิติ Z และ F เท่านั้น

1.5.3 ขนาดตัวอย่างที่ใช้ศึกษาจะกำหนดให้ทุกกลุ่มตัวอย่างมีขนาด 30, 50 และ100

1.5.4 กำหนดนัยสำคัญ 0.05 และ 0.10

1.5.5 การทดสอบเมื่อใช้ตัวสถิติ Kolmogorov-Smirnov Test จะกำหนดพารามิเตอร์ β ของการแจกแจงแบบเอกซ์โปเนนเชียลเท่ากับค่าเฉลี่ย ($\bar{x}$) ซึ่งเป็นตัวประมาณแบบ moment และตัวประมาณแบบ maximum likelihood ของพารามิเตอร์ β

1.5.6 กรณีมีข้อมูลขาดหายในการวิเคราะห์จะพิจารณาจำนวนข้อมูลขาดหายทางซ้าย (left censored data = r_1) และ จำนวนข้อมูลขาดหายทางขวา (right censored data = r_2) โดยพิจารณาจากกรณีต่อไปนี้

กรณีที่ 1 การวิเคราะห์ข้อมูลทั้งหมดจะกำหนดให้ $r_1 = 0$ และ $r_2 = 0$

กรณีที่ 2 การวิเคราะห์ข้อมูลเมื่อมีข้อมูลขาดหายทางขวาจะกำหนดให้ $r_1 = 0$ และ r_2 = 10%,20% ของจำนวนข้อมูล

กรณีที่ 3 การวิเคราะห์ข้อมูลเมื่อมีข้อมูลขาดหายทางซ้ายจะกำหนดให้ $r_2 = 0$ และ r_1 = 10%,20% ของจำนวนข้อมูล

กรณีที่ 4 การวิเคราะห์ข้อมูลเมื่อมีข้อมูลขาดหายทางซ้ายและทางขวา จะกำหนดให้ $r_1 = r_2 = r$ โดยที่ r เท่ากับ 10% และ 20% ของจำนวนข้อมูลตามลำดับ

1.5.7 การวิจัยครั้งนี้จะสร้างแบบจำลองข้อมูลให้มีสถานการณ์ตามต้องการศึกษาโดยใช้เทคนิคมอนติคาร์โลซิมูเลชันจากเครื่องคอมพิวเตอร์ IBM 370/3031 โดยเขียนโปรแกรมด้วยภาษาฟอร์แทรน 77 และทำการทดลองซ้ำประมาณ 500 ครั้งในแต่ละสถานการณ์

1.6 คำจำกัดความ

ก. อำนาจการทดสอบ (power of test) หมายถึง ความน่าจะเป็นที่จะปฏิเสธสมมุติฐานว่าง (H_0) เมื่อสมมุติฐานว่างไม่จริง

ข. ความคลาดเคลื่อนประเภทที่ 1 (type 1 error) หมายถึง ความคลาดเคลื่อนที่เกิดจากการปฏิเสธสมมุติฐานว่างเมื่อสมมุติฐานว่างจริง

1.7 ประโยชน์ที่คาดว่าจะได้รับ

เพื่อเป็นแนวทางให้ผู้วิจัยได้เลือกใช้ตัวสถิติทดสอบที่เหมาะสมสำหรับการทดสอบการแจกแจงแบบเอกซ์โปเนนเชียลในกรณีที่ไม่ทราบแน่ชัดเกี่ยวกับลักษณะการแจกแจงของประชากร และในกรณีที่เกิดข้อมูลขาดหายในการวิเคราะห์