**ใบงานที่ 3**

1. จากข้อมูลการวางแผนครอบครัวแสดงดังตารางที่ 1 ข้างล่างนี้ และมีคำอธิบายตัวแปรดังตารางที่ 2 ให้นักศึกษาอธิบายการทำความสะอาดข้อมูลของแต่ละตัวแปร พร้อมเขียนคำสั่งในโปรแกรม R

**ตารางที่ 1** ข้อมูลการวางแผนครอบครัว



**ตารางที่ 2** คำอธิบายตัวแปรของข้อมูลการวางแผนครอบครัว



1. ตัวแปร AGE

ปัญหาที่เจอ และวิธีการจัดการข้อมูล

ปัญหาที่เจอคือ ตัวแปร AGE มีข้อมูลสูญหายและอายุ 8ปี ซึ่งไม่สอดคล้องกับความเป็นจริง

วิธีการจัดการข้อมูล คือ ต้องเปลี่ยนข้อมูลทั้งหมดมาเป็น NA

คำสั่งในโปรแกรม R

dt$AGE <- ifelse(dt$AGE==”8” | dt$AGE==” ”,NA, dt$AGE)

1. ตัวแปร RELIG

ปัญหาที่เจอ และวิธีการจัดการข้อมูล

ปัญหาที่เจอคือ ตัวแปร RELIG มีข้อมูลศาสนาเป็น 9 และ -1 ซึ่งไม่สอดคล้องกับความเป็นจริง

วิธีการจัดการข้อมูล คือ ต้องเปลี่ยนข้อมูลที่ไม่สอดคล้องมาเป็น NA

คำสั่งในโปรแกรม R

dt$RELIG<- ifelse(dt$RELIG== -1 | dt$RELIG== 9,NA, dt$RELIG)

1. ตัวแปร PED

ปัญหาที่เจอ และวิธีการจัดการข้อมูล

ปัญหาที่เจอคือ ตัวแปร PED มีข้อมูลระดับการศึกษา เป็น 9 ซึ่งไม่สอดคล้องกับข้อมูล

วิธีการจัดการข้อมูล คือ ต้องเปลี่ยนข้อมูลที่ไม่สอดคล้องมาเป็น NA

คำสั่งในโปรแกรม R

dt$PED <- ifelse(dt$PED==9,NA,dt$PED)

1. ตัวแปร INCOME

ปัญหาที่เจอ และวิธีการจัดการข้อมูล

ปัญหาที่เจอคือ ตัวแปร INCOME มีข้อมูลเงินเดือน เป็น 9 ซึ่งไม่สอดคล้องกับความเป็นจริง

วิธีการจัดการข้อมูล คือ ต้องเปลี่ยนข้อมูลที่ไม่สอดคล้องมาเป็น NA

คำสั่งในโปรแกรม R

dt$INCOME <- ifelse(dt$INCOME==9,NA,dt$INCOME)

1. ตัวแปร AM

ปัญหาที่เจอ และวิธีการจัดการข้อมูล

ไม่มีข้อมูลที่ผิด

คำสั่งในโปรแกรม R

……………………………………………………………………………………………………………………………………………………………..

……………………………………………………………………………………………………………………………………………………………..

1. ตัวแปร REASON

ปัญหาที่เจอ และวิธีการจัดการข้อมูล

ไม่มีข้อมูลที่ผิด

คำสั่งในโปรแกรม R

……………………………………………………………………………………………………………………………………………………………..

……………………………………………………………………………………………………………………………………………………………..

1. ตัวแปร BPS

ปัญหาที่เจอ และวิธีการจัดการข้อมูล

ปัญหาที่เจอคือ ตัวแปร BPS มีข้อมูลความดันโลหิตBPS เป็น999 ซึ่งไม่สอดคล้องกับความเป็นจริง

วิธีการจัดการข้อมูล คือ ต้องเปลี่ยนข้อมูลที่ไม่สอดคล้องมาเป็น NA

คำสั่งในโปรแกรม R

dt$BPS <- ifelse(dt$BPS==999,NA,dt$BPS)

1. ตัวแปร BPD

ปัญหาที่เจอ และวิธีการจัดการข้อมูล

ปัญหาที่เจอคือ ตัวแปร BPD มีข้อมูลความดันโลหิตBPD เป็น999 ซึ่งไม่สอดคล้องกับความเป็นจริง

วิธีการจัดการข้อมูล คือ ต้องเปลี่ยนข้อมูลที่ไม่สอดคล้องมาเป็น NA

คำสั่งในโปรแกรม R

dt$BPD <- ifelse(dt$BPD==999,NA,dt$BPD)

1. ตัวแปร WT

ปัญหาที่เจอ และวิธีการจัดการข้อมูล

ปัญหาที่เจอคือ ตัวแปร WT มีน้ำหนัก เป็น1150 ซึ่งไม่สอดคล้องกับความเป็นจริง

วิธีการจัดการข้อมูล คือ ต้องเปลี่ยนข้อมูลที่ไม่สอดคล้องมาเป็น NA

คำสั่งในโปรแกรม R

dt$WT <- ifelse(dt$WT==150,NA,dt$WT)

1. ตัวแปร HT

ปัญหาที่เจอ และวิธีการจัดการข้อมูล

ปัญหาที่เจอคือ ตัวแปร BPD มีส่วนสูง เป็น999 และ1520 ซึ่งไม่สอดคล้องกับความเป็นจริง

วิธีการจัดการข้อมูล คือ ต้องเปลี่ยนข้อมูลที่ไม่สอดคล้องมาเป็น NA

คำสั่งในโปรแกรม R

dt$HT<- ifelse(dt$HT==999 | dt$HT == 1520,NA,dt$HT)