

Homework 5
Zhaoxing Wu
Stat 443

- ADJINC: This variable is adjustment factor for income and earning dollar amounts, and this dataset all has the value of 1010145 and cannot help out estimation of INTP.
- DECADE: this variable is decade of entry. There is another YPEP, year of entry. YPEP is more detailed than DECADE, so we need to delete DECADE to avoid dependency.
- DIS: this variable is disability recode, whose value is either with disability or without disability. The variables DDRS, DEAR, DEYE, DOUT, DPHY, DRAT, DRATX, DREM are about different certain types of disability, and DIS is like a summary of all kinds of disability. Therefore, we need to delete this to avoid dependency.
- DIVISION: This variable is the division code and there are 10 in total. Since the data are from Minnesota, they are all decoded 4 as "west North central" and cannot help our estimation of INTP.
- DRIVESP: this variable is the number of vehicles calculated from JWRI, which is vehicle occupancy. We need to delete this to avoid linear dependency.
- FDISP: this variable is disability recode allocation flag. Since DIS should be deleted, the corresponding allocation flag should be deleted as well.
- FHICOVP: this variable is health insurance coverage recode allocation flag. Since HICOV should be deleted, the corresponding allocation flag should be deleted as well.
- FLANXP: this variable is language other than English allocation flag. Since LANX should be deleted, the corresponding allocation flag should be deleted as well.
- FMARP: this variable is marital status allocation flag. Since MAR should be deleted, the corresponding allocation flag should be deleted as well.
- FMIGP: this variable is mobility status allocation flag. Since MIG should be deleted, the corresponding allocation flag should be deleted as well.
- FMILSP: this variable is military service allocation flag. Since MIL should be deleted, the corresponding allocation flag should be deleted as well.
- FOCCP: this variable is occupation allocation flag. Since OCCP should be deleted, the corresponding allocation flag should be deleted as well.
- HICOV: This variable is health insurance coverage recode, and it can be either with health insurance coverage or no health insurance coverage. The variables HINS1, HINS2, HINS3, HINS4, HINS5, HINS6, HINS7 have define whether the person have certain kinds of health insurance, so we need to delete this variable because it is linearly dependent on the other variables mentioned above.
- LANX: This variable is language other than English spoken at home. The variable LANP specifies which language spoken at home. It contained the information of LANX and more than that, so LANX should be deleted to avoid dependency.
- MAR: This variable is marital status, and there is another variable MSP, which is married, spouse present/spouse absent. MSP not only contains the same information as MAR, and it also contains more information of the presence of spouse, so it is necessary to keep MSP and delete MAR.

MIG: This variable is Mobility status (lived here 1 year ago). The variable VPS, veteran period of service, contains the information of this variable, so delete this to avoid dependency.

MIL: This variable is Military service. The variable VPS, veteran period of service, contains the information of this variable, so delete this to avoid dependency.

MLPA: This variable is Served September 2001 or later. The variable VPS, veteran period of service, contains the information of this variable, so delete this to avoid dependency.

MLPB: This variable is Served August 1990 - August 2001 (including Persian Gulf War). The variable VPS, veteran period of service, contains the information of this variable, so delete this to avoid dependency.

MLPCD: This variable is Served May 1975 - July 1990. The variable VPS, veteran period of service, contains the information of this variable, so delete this to avoid dependency.

MLPE: This variable is Served Vietnam era (August 1964 - April 1975). The variable VPS, veteran period of service, contains the information of this variable, so delete this to avoid dependency.

MLPFG: This variable is Served February 1955 - July 1964. The variable VPS, veteran period of service, contains the information of this variable, so delete this to avoid dependency.

MLPH: This variable is Served Korean War (July 1950 - January 1955). The variable VPS, veteran period of service, contains the information of this variable, so delete this to avoid dependency.

MLPI: This variable is Served January 1947 - June 1950. The variable VPS, veteran period of service, contains the information of this variable, so delete this to avoid dependency.

MLPJ: This variable is Served World War II (December 1941 - December 1946). The variable VPS, veteran period of service, contains the information of this variable, so delete this to avoid dependency.

MLPK: This variable is Served November 1941 or earlier. The variable VPS, veteran period of service, contains the information of this variable, so delete this to avoid dependency.

OCCP: this variable is occupation recode for 2018 and later based on 2018 OCC codes. The variable SOCP is Standard Occupational Classification (SOC) codes for 2018 and later based on 2018 SOC codes. They contain roughly the same information about occupation, but SOCP is more detailed, so delete OCCP to avoid dependency.

PWGTP1-80: These 80 variables are person's weight replicate 1-80, which is used to calculate confidence interval. These should be deleted since it does not help the estimation of INTP.

RAC1P: This variable is recoded detailed race code. The variable RAC3P is more detailed than this one, so keeping RAC3P and deleting RAC1P to avoid dependency.

RAC2P: This variable is recoded detailed race code. The variable RAC3P is more detailed than this one, so keeping RAC3P and deleting RAC2P to avoid dependency.

RACAIAN: This variable is American Indian, and Alaska Native recode (American Indian and Alaska Native alone or in combination with one or more other races). The variable RAC3P already contains the information in this variable, so deleting to avoid dependency.

RACASN: This variable is Asian recode (Asian alone or in combination with one or more other races). The variable RAC3P already contains the information in this variable, so deleting to avoid dependency.

RACBLK: This variable is Black or African American recode (Black alone or in combination with one or more other races). The variable RAC3P already contains the information in this variable, so deleting to avoid dependency.

RACNH: This variable is Native Hawaiian recode (Native Hawaiian alone or in combination with one or more other races). The variable RAC3P already contains the information in this variable, so deleting to avoid dependency.

RACPI: This variable is Other Pacific Islander recode (Other Pacific Islander alone or in combination with one or more other races). The variable RAC3P already contains the information in this variable, so deleting to avoid dependency.

RACSOR: This variable is Some other races recode (Some other race alone or in combination with one or more other races). The variable RAC3P already contains the information in this variable, so deleting to avoid dependency.

RACWHT: This variable is White recode (White alone or in combination with one or more other races). The variable RAC3P already contains the information in this variable, so deleting to avoid dependency.

REGION: This variable is region code and there are 5 in total. Since the data are from Minnesota, they are all decoded 2 as “Midwest” and cannot help our estimation of INTP.

RT: This variable is record type, because this dataset is about person record, so this variable is the same for each row, and cannot help our estimation of INTP.

SERIALNO: this variable is GQ person serial number, which is used with SPORDER as a unique identifier for each person. This is completely irrelevant to INTP and should be deleted.

SPORDER: this variable is person number, which is used with SERIALNO as a unique identifier for each person. This is completely irrelevant to INTP and should be deleted.

ST: This variable is state code and there are 52 in total. Since the data are from Minnesota, they are all decoded 27 as “Minnesota/MN” and cannot help our estimation of INTP.