

```

capture log close
log using "missing_data", replace text

set more off
set linesize 80
set scheme slmono
clear all
macro drop _all

// Program: Missing Data.do
// Task: Sample do file for missing data simulation
// Author: Angran Li
// Note: This Do-file simulates missing data mechanisms (MCAR, MAR, and NMAR).
// Please set up working directory before running this do-file.
// Please contact me at angran.li@uconn.edu for more information.

* Missing Data Simulation

// Program description:
// ==> In this do file, I conduct Monte Carlo simulations for detecting how missing data
// ==> mechanisms affect OLS regression estimators. It provides some evidence to help
// ==> students to understand the potential bias when data is not missing at random,
// ==> which is likely to occur for studies using observational data. For data
simulation,
// ==> I create random continues and binary variables to simulate the results.

// No missing data

clear all
program testregress
    drop _all
    set obs 1000
    gen x2 = rnormal()
    gen x3 = runiform()
    replace x3=1 if x3>=.5
    replace x3=0 if x3<.5
    gen x1 = .5*x2+.5*x3+rnormal()
    gen y = 10+2*x1+3*x2+4*x3+rnormal()
    regress y x1 x2 x3
        test x1=2
        gen p1 = r(p)
        test x2=3
        gen p2 = r(p)
        test x3=4
        gen p3 = r(p)
end

simulate _b _se p1 p2 p3, reps(1000): testregress

sum _b_x1 _b_x2 _b_x3 _b_cons

gen reject_x1 =0
replace reject_x1=1 if _eq2_sim_1<.05

```

```

gen reject_x2 =0
replace reject_x2=1 if _eq2_sim_2<.05

gen reject_x3 =0
replace reject_x3=1 if _eq2_sim_3<.05

tab reject_x1
tab reject_x2
tab reject_x3

// Missing completely at random (MCAR)

clear all
program MCAR
    drop _all
    set obs 1000
    gen x2 = rnormal()
    gen x3 = runiform()
    replace x3=1 if x3>=.5
    replace x3=0 if x3<.5
    gen x1 = .5*x2+.5*x3+rnormal()
    gen y = 10+2*x1+3*x2+4*x3+rnormal()
    randomtag, count(100) gen(mx1)
    randomtag, count(100) gen(mx2)
        randomtag, count(100) gen(mx3)
        randomtag, count(100) gen(my)
        replace x1=. if mx1==1
replace x2=. if mx2==1
replace x3=. if mx3==1
replace y=. if my==1
regress y x1 x2 x3
    test x1=2
    gen p1 = r(p)
    test x2=3
    gen p2 = r(p)
    test x3=4
    gen p3 = r(p)
end

simulate _b _se p1 p2 p3, reps(1000): MCAR

sum _b_x1 _b_x2 _b_x3 _b_cons

gen reject_x1 =0
replace reject_x1=1 if _eq2_sim_1<.05

gen reject_x2 =0
replace reject_x2=1 if _eq2_sim_2<.05

gen reject_x3 =0
replace reject_x3=1 if _eq2_sim_3<.05

```

```

tab reject_x1
tab reject_x2
tab reject_x3

// Missing at random (MAR)

clear all
program MAR
    drop _all
    set obs 1000
    gen x2 = rnormal()
    gen x3 = runiform()
    replace x3=1 if x3>=.5
    replace x3=0 if x3<.5
    gen x1 = .5*x2+.5*x3+rnormal()
    gen y = 10+2*x1+3*x2+4*x3+rnormal()
    replace x1=. if x2<-1
    replace y=. if x1<-1
    regress y x1 x2 x3
    test x1=2
    gen p1 = r(p)
    test x2=3
    gen p2 = r(p)
    test x3=4
    gen p3 = r(p)
end

simulate _b _se p1 p2 p3, reps(1000): MAR

sum _b_x1 _b_x2 _b_x3 _b_cons

gen reject_x1 =0
replace reject_x1=1 if _eq2_sim_1<.05

gen reject_x2 =0
replace reject_x2=1 if _eq2_sim_2<.05

gen reject_x3 =0
replace reject_x3=1 if _eq2_sim_3<.05

tab reject_x1
tab reject_x2
tab reject_x3

// Not missing at random (NMAR)

clear all
program NMAR
    drop _all
    set obs 1000
    gen x2 = rnormal()
    gen x3 = runiform()
    replace x3=1 if x3>=.5

```

```

replace x3=0 if x3<.5
gen x1 = .5*x2+.5*x3+rnormal()
gen y = 10+2*x1+3*x2+4*x3+rnormal()
replace y=. if y<15
regress y x1 x2 x3
    test x1=2
    gen p1 = r(p)
    test x2=3
    gen p2 = r(p)
    test x3=4
    gen p3 = r(p)
end

simulate _b _se p1 p2 p3, reps(1000): NMAR

sum _b_x1 _b_x2 _b_x3 _b_cons

gen reject_x1 =0
replace reject_x1=1 if _eq2_sim_1<.05

gen reject_x2 =0
replace reject_x2=1 if _eq2_sim_2<.05

gen reject_x3 =0
replace reject_x3=1 if _eq2_sim_3<.05

tab reject_x1
tab reject_x2
tab reject_x3

log close
exit

```