

```

/* draw a random sample of size 300,000 */

set seed 33948
sample 300000, count

/* MCAR */
foreach var of varlist hihr lsys htemp lrr pf loph hul hicreat1 hinal lwbc1
dup spgcs {
generate `var' _i=`var'
replace `var' _i=. if runiform()<.02
}

```

Variables examined for missing values

#	Variable	# Missing	% Missing
1	hihr_i	5965	2.0
2	lsys_i	5940	2.0
3	htemp_i	5957	2.0
4	lrr_i	6128	2.0
5	pf_i	5922	2.0
6	loph_i	5974	2.0
7	hul_i	5947	2.0
8	hicreat1_i	5931	2.0
9	hinal_i	6050	2.0
10	lwbc1_i	6008	2.0
11	dup_i	6025	2.0
12	spgcs_i	5894	2.0

Missing for how many variables?	Freq.	Percent	Cum.
0	235,572	78.52	78.52
1	57,603	19.20	97.72
2	6,359	2.12	99.84
3	444	0.15	99.99
4	22	0.01	100.00
Total	300,000	100.0	

```

/* MAR */
gen double missing1=3-
2*loph+0.7*ahsurv+1.5*vent1*low_risk+0.1*yulos+0.05*calage+1*short_free
gen prb_missing1=invlogit(missing1)
foreach var of varlist hihr lsys htemp lrr {
generate `var' _i=`var'
replace `var' _i=. if prb_missing1<=0.0009
}
gen double missing2=3-0.3*ahsurv+1*vent-1*low_risk+ 0.05*yulos-
0.07*calage+1*short_free
gen prb_missing2=invlogit(missing2)
foreach var of varlist pf loph {
generate `var' _i=`var'
replace `var' _i=. if prb_missing2<=0.2
}
gen double missing3=3-2*loph+0.7*ahsurv+1.5*vent -1*low_risk+
0.1*yulos+0.05*calage
gen prb_missing3=invlogit(missing3)
foreach var of varlist hul hicreat1 hinal lwbc1 {
generate `var' _i=`var'
replace `var' _i=. if prb_missing3<=0.0009
}

```

```

gen double missing4=3-2*loph+0.7*ahsurv+1.5*vent -1*low_risk+
0.1*yulos+0.05*calage
gen prb_missing4=invlogit(missing4)
gen dup_i=dup
replace dup_i=. if prb_missing4<=0.0009

gen double missing5=3-2*loph+0.7*ahsurv+1.5*vent -1*low_risk+
0.1*yulos+0.05*calage
gen prb_missing5=invlogit(missing5)
gen spgcs_i=spgcs
replace spgcs_i=. if prb_missing5<=0.0009

```

Variables examined for missing values

#	Variable	# Missing	% Missing
1	hihr_i	36537	12.2
2	lsys_i	36537	12.2
3	htemp_i	36537	12.2
4	lrr_i	36537	12.2
5	pf_i	34800	11.6
6	loph_i	34800	11.6
7	hul_i	39085	13.0
8	hicreat1_i	39085	13.0
9	hinal_i	39085	13.0
10	lwbc1_i	39085	13.0
11	dup_i	39085	13.0
12	spgcs_i	39085	13.0

Missing for |
how many |
variables? |

	Freq.	Percent	Cum.
0	238,013	79.34	79.34
2	22,902	7.63	86.97
6	2,492	0.83	87.80
8	56	0.02	87.82
10	24,695	8.23	96.05
12	11,842	3.95	100.00
Total	300,000	100.00	