

```

-----
> -----
      name: <unnamed>
      log: C:\Users\saiomkark\OneDrive\Documents\AdvStats\anRoyModelSample.log
      log type: text
      opened on: 8 Oct 2021, 19:09:38

. *
. * Set up data simulation dataset. Set seed so we may replicate this program
. *
.
. *****
> *****
.
.
. set seed 195912191

. *
. *
. * Set covariance matrix for Standard normal random variables
. *
. matrix V = (1, .9 \ .9, 1)

. * draw bivariate normal with n = 1,000,000
. drawnorm za ze, cov(V) n(1000000)
(obs 1,000,000)

. *
. * y1 economists y0 accountant mean 60k for econ, 65k for acct
. * sd econ 10,000 and 5,000 for acct
. *
. gen y1 = 60000 + 10000*ze

.
.
. gen y0 = 65000 + 5000*za

. *summarize the data
. sum y1 y0

      Variable |           Obs       Mean   Std. Dev.       Min       Max
-----+-----
      y1 | 1,000,000   60004.08   9995.004   12692.68   105929
      y0 | 1,000,000   65001.36   4999.691   40285.02   89629.73

. * pick the field that pays the most
. gen y = max(y0,y1)

. * summarize observed earnings
. sum y

      Variable |           Obs       Mean   Std. Dev.       Min       Max
-----+-----
      y | 1,000,000   65656.41   6016.329   40285.02   105929

. * d=1 means economics, d=0 accounting
. gen d = y1 > y0

. * run cross tables
. tab d

      d |           Freq.       Percent       Cum.
-----+-----
      0 |   801,660       80.17       80.17
      1 |   198,340       19.83      100.00
-----+-----
      Total | 1,000,000      100.00

```

```
. tab d, sum(y)
```

d		Summary of y		Freq.
		Mean	Std. Dev.	
0		63826.404	4499.5368	801,660
1		73053.032	5693.8071	198,340
Total		65656.413	6016.3287	1,000,000

```
. * selection into y0
```

```
. tab d, sum(y0)
```

d		Summary of y0		Freq.
		Mean	Std. Dev.	
0		63826.404	4499.5368	801,660
1		69750.352	4008.4031	198,340
Total		65001.36	4999.6907	1,000,000

```
. * selection into y1
```

```
. tab d, sum(y1)
```

d		Summary of y1		Freq.
		Mean	Std. Dev.	
0		56775.624	8002.7879	801,660
1		73053.032	5693.8071	198,340
Total		60004.085	9995.0037	1,000,000

```
.
. *
.
. *Answer for Question1
. *Two reasons:
. *Seed is different in R and Stata. Seed is a dynamic number that can be set which al
> lows us to replicate the observations generated b
> y the random number generator. It is helpful for simulations to replicate the result
> s which can aid simulation analysis.
. *However, even with same seed, different results are observed in R and Stata. This c
> an be because the Normal distribution generated i
> n R differs from what is generated in Stata.
. *This shows that the underlying distribution generator algorithm functions different
> ly in R(mvrnorm) and Stata(drawnorm).
. *
.
.
.
. *****
.
. * makes sure memory is clear
. clear
.
.
. *Question2
.   *Change the seed
.   *set seed 02101870
. *Question2
```

```

.
.
. set seed 195912191

.
. *
. * Set covariance matrix for Standard normal random variables
. *
. matrix V = (1, .9 \ .9, 1)

. * draw bivariate normal with n = 1,000,000
. drawnorm za ze, cov(V) n(1000000)
(obs 1,000,000)

. *
. * y1 economists y0 accountant mean 60k for econ, 65k for acct
. * sd econ 10,000 and 5,000 for acct
. *
. gen y1 = 60000 + 10000*ze

.
.
. gen y0 = 65000 + 5000*za

. *summarize the data
. sum y1 y0

```

Variable	Obs	Mean	Std. Dev.	Min	Max
y1	1,000,000	60004.08	9995.004	12692.68	105929
y0	1,000,000	65001.36	4999.691	40285.02	89629.73

```

. * pick the field that pays the most
. gen y = max(y0,y1)

. * summarize observed earnings
. sum y

```

Variable	Obs	Mean	Std. Dev.	Min	Max
y	1,000,000	65656.41	6016.329	40285.02	105929

```

. * d=1 means economics, d=0 accounting
. gen d = y1 > y0

. * run cross tables
. tab d

```

d	Freq.	Percent	Cum.
0	801,660	80.17	80.17
1	198,340	19.83	100.00
Total	1,000,000	100.00	

```

. tab d, sum(y)

```

d	Summary of y		Freq.
	Mean	Std. Dev.	
0	63826.404	4499.5368	801,660
1	73053.032	5693.8071	198,340
Total	65656.413	6016.3287	1,000,000

```
. * selection into y0
. tab d, sum(y0)
```

d		Summary of y0		Freq.
		Mean	Std. Dev.	
0		63826.404	4499.5368	801,660
1		69750.352	4008.4031	198,340
Total		65001.36	4999.6907	1,000,000

```
. * selection into y1
. tab d, sum(y1)
```

d		Summary of y1		Freq.
		Mean	Std. Dev.	
0		56775.624	8002.7879	801,660
1		73053.032	5693.8071	198,340
Total		60004.085	9995.0037	1,000,000

```
.
.
.
. *Answer for Question 2
. *Reset the seed to original seed value 195912191
. *1) Reason for not giving birthdate of self could be construed a data privacy issue
> or data security issue. Also it will produce similar results as fellow mates in the class because the birthday paradox shows that the
> probability is a greater probability for repetition of birthday
> any when the sample size exceeds 23.
. *2) Seed set is different in question 1 and question 2. As seed is changed, the distribution
> generated changes slightly causing the results to change slightly.
. *However the mean and variance still remains same as they are fixed. Only the
> observations change with change in seed.
.
.
.
.
.
.
.
. *****
. * makes sure memory is clear
. clear
.
.
. set seed 195912191
.
.
. *
. * Set covariance matrix for Standard normal random variables
. *
. matrix V = (1, .9 \ .9, 1)

. * draw bivariate normal with n = 1,000,000
. drawnorm za ze, cov(V) n(1000000)
(obs 1,000,000)
```

```

. *
. * y1 economists y0 accountant mean 60k for econ, 65k for acct
. * sd econ 10,000 and 5,000 for acct
. *
. gen y1 = 60000 + 10000*ze

.
. gen y0 = 65000 + 5000*za

. *summarize the data
. sum y1 y0

```

Variable	Obs	Mean	Std. Dev.	Min	Max
y1	1,000,000	60004.08	9995.004	12692.68	105929
y0	1,000,000	65001.36	4999.691	40285.02	89629.73

```

. * pick the field that pays the most
. gen y = max(y0,y1)

. * summarize observed earnings
. sum y

```

Variable	Obs	Mean	Std. Dev.	Min	Max
y	1,000,000	65656.41	6016.329	40285.02	105929

```

. * d=1 means economics, d=0 accounting
. gen d = y1 > y0

. * run cross tables
. tab d

```

d	Freq.	Percent	Cum.
0	801,660	80.17	80.17
1	198,340	19.83	100.00
Total	1,000,000	100.00	

```

. tab d, sum(y)

```

d	Summary of y		
	Mean	Std. Dev.	Freq.
0	63826.404	4499.5368	801,660
1	73053.032	5693.8071	198,340
Total	65656.413	6016.3287	1,000,000

```

. * selection into y0
. tab d, sum(y0)

```

d	Summary of y0		
	Mean	Std. Dev.	Freq.
0	63826.404	4499.5368	801,660
1	69750.352	4008.4031	198,340
Total	65001.36	4999.6907	1,000,000

```
. * selection into y1
. tab d, sum(y1)
```

d	Summary of y1		
	Mean	Std. Dev.	Freq.
0	56775.624	8002.7879	801,660
1	73053.032	5693.8071	198,340
Total	60004.085	9995.0037	1,000,000

```
.
. *
.
. **Answer for Question 3
. *The standard normal sample size is 1000000
. *It is observed that out of the sample size for economists only 198340 have become e
> conomists and hence the standard deviation we obs
> erved (5693.8071) is significantly less compared to the conditional standard deviati
> on of 100000
. *The same goes for accountants, out of the sample size for accountants its observed
> that 801660 have become accountants, hence the st
> andard deviation we observed (4499.5368) is less compared to the conditional standar
> d deviation set of 5000
.
.
.
.
. *****
.
. * makes sure memory is clear
. clear
.
. *Question4
. *Change the standard deviation of Economist earnings to 12000 from 10000
. * gen y1 = 60000 + 12000*ze
. *Question4
.
.
. set seed 195912191
.
.
. *
. * Set covariance matrix for Standard normal random variables
. *
. matrix V = (1, .9 \ .9, 1)
.
. * draw bivariate normal with n = 1,000,000
. drawnorm za ze, cov(V) n(1000000)
(obs 1,000,000)
.
. *
. * y1 economists y0 accountant mean 60k for econ, 65k for acct
. * sd econ 10,000 and 5,000 for acct
. *
. gen y1 = 60000 + 12000*ze
.
.
```

```
. gen y0 = 65000 + 5000*za
```

```
. *summarize the data
```

```
. sum y1 y0
```

Variable	Obs	Mean	Std. Dev.	Min	Max
y1	1,000,000	60004.9	11994	3231.216	115114.7
y0	1,000,000	65001.36	4999.691	40285.02	89629.73

```
. * pick the field that pays the most
```

```
. gen y = max(y0,y1)
```

```
. * summarize observed earnings
```

```
. sum y
```

Variable	Obs	Mean	Std. Dev.	Min	Max
y	1,000,000	66232.43	6952.761	40285.02	115114.7

```
. * d=1 means economics, d=0 accounting
```

```
. gen d = y1 > y0
```

```
. * run cross tables
```

```
. tab d
```

d	Freq.	Percent	Cum.
0	739,328	73.93	73.93
1	260,672	26.07	100.00
Total	1,000,000	100.00	

```
. tab d, sum(y)
```

d	Mean	Std. Dev.	Freq.
0	63369.132	4294.5497	739,328
1	74353.441	6628.3023	260,672
Total	66232.434	6952.7611	1,000,000

```
. * selection into y0
```

```
. tab d, sum(y0)
```

d	Mean	Std. Dev.	Freq.
0	63369.132	4294.5497	739,328
1	69630.746	3820.6677	260,672
Total	65001.36	4999.6907	1,000,000

```
. * selection into y1
```

```
. tab d, sum(y1)
```

d	Mean	Std. Dev.	Freq.
0	54945.899	8994.6477	739,328
1	74353.441	6628.3023	260,672
Total	60004.902	11994.004	1,000,000

```

.
. *Answer to question 4
. *The data shows an increase in total economists and decrease in accountants. As the
> standard deviation of economist earnings is incre
> ased from $10000 to $12000, the economist earnings of economist have increased and a
> ccountants with economist earnings have reduced.
> This is due to the flattening of the curve that happens because of the increase in s
> tandard deviation causing the occurrences of high
> economist earnings has increased resulting more total economists and less total acc
> ountants. Due to the same, the mean of economists
> with economist earnings increases (Mean changed from 73053.0.32 to 74353.441 ) whe
> reas the mean of accountants with accountant earn
> ings has reduced(Mean changed from 63826.404 to 63369.132). One another observation
> is the increase in people with economists earning
> s (198340 to 260672) and reduction in number of people with accountants earnings (80
> 1660 to 739328)
.
.
. *****
.
.
. *
. log close
    name: <unnamed>
    log: C:\Users\saiomkark\OneDrive\Documents\AdvStats\anRoyModelSample.log
    log type: text
    closed on: 8 Oct 2021, 19:09:42
-----
> -----

```