

```

-----
      name: <unnamed>
      log: C:\Users\saiomkark\OneDrive - The University of Chicago\AdvStats\
> PS5\Question2.log
      log type: text
opened on: 5 Nov 2021, 19:31:36

.
. clear

. *For this exercise, you will use a simulation to see how well the CLT works
> with finite samples in R. Parts (a) and (b) of this question each describe a
> distribution. For each distribution, use 10,000 draws using each of the fol
> lowing sample sizes: n = 36, n = 64, n = 100, n = 225, n = 2500, and n = 121
> 00. Then discuss how well the normal approximation fits your simulated estim
> ates of the means at the critical values of 0.025 and 0.975.
. * a) Suppose that x is binary with  $\Pr(x = 1) = 0.35$ .
. * b) Suppose that x is binary with  $\Pr(x = 1) = 0.97$ 
.
.
. capture program drop clt

. program clt, rclass
1. version 15.0
2.
.
.
.
. args N
3. clear
4.
. *
. *
. quietly set obs `N'
5. *
. *
. tempvar y1 y2
6. *
. * y1 is binary with  $\Pr(y1=1) = 0.35$ 
. *
. gen `y1' = uniform() < 0.35
7. *
. * y2 is binary with  $\Pr(y1=1) = 0.97$ 
. *
. gen `y2' = uniform() < 0.97
8. *
. *
. quietly sum `y1'
9. *
. *
. return scalar mu1 = r(mean)
10. return scalar sd1 = r(sd)
11. *
. *
. quietly sum `y2'
12. *
. *
. return scalar mu2 = r(mean)
13. return scalar sd2 = r(sd)
14. *

```

```

. *
. end

. *
. *
. set seed 24031997

. *
. *
. simulate mu1 = r(mu1) mu2 = r(mu2) sd1 = r(sd1) sd2 = r(sd2), reps(10000) no
> dots: clt 36

```

```

      command:  clt 36
      mu1:      r(mu1)
      mu2:      r(mu2)
      sd1:      r(sd1)
      sd2:      r(sd2)

```

```

. *
. *
. save clt36, replace
file clt36.dta saved

```

```

. *
. * Sample of 36 observations
. *
. use clt36
(simulate: clt)

```

```

. gen z1 = (mu1-0.35)          /(sd1/36^(1/2))
. gen z2 = (mu2-0.97)          /(sd2/36^(1/2))
(3,360 missing values generated)

```

```

. *
. *
. gen rrej1 = z1 > 1.96 & z1 < .
. gen lrej1 = z1 < -1.96
. gen rrej2 = z2 > 1.96
. gen lrej2 = z2 < -1.96

. *
. *
. sum rrej1-lrej2

```

Variable	Obs	Mean	Std. Dev.	Min	Max
rrej1	10,000	.0208	.1427214	0	1
lrej1	10,000	.0307	.1725123	0	1
rrej2	10,000	.336	.4723625	0	1
lrej2	10,000	.0005	.0223562	0	1

```

.
. bitest lrej1 = 0.025

```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej1	10000	307	250	0.02500	0.03070

```

Pr(k >= 307)          = 0.000227 (one-sided test)
Pr(k <= 307)          = 0.999819 (one-sided test)
Pr(k <= 196 or k >= 307) = 0.000421 (two-sided test)

```

```
. bitest rrej1 = 0.025
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej1	10000	208	250	0.02500	0.02080

```
Pr(k >= 208) = 0.997387 (one-sided test)
Pr(k <= 208) = 0.003217 (one-sided test)
Pr(k <= 208 or k >= 294) = 0.006455 (two-sided test)
```

```
. bitest lrej2 = 0.025
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej2	10000	5	250	0.02500	0.00050

```
Pr(k >= 5) = 1.000000 (one-sided test)
Pr(k <= 5) = 0.000000 (one-sided test)
Pr(k <= 5 or k >= 647) = 0.000000 (two-sided test)
```

```
. bitest rrej2 = 0.025
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej2	10000	3360	250	0.02500	0.33600

```
Pr(k >= 3360) = 0.000000 (one-sided test)
Pr(k <= 3360) = 1.000000 (one-sided test)
Pr(k >= 3360) = 0.000000 (two-sided test)
```

note: lower tail of two-sided p-value is empty

```
.
.
. replace rrej1 = z1 > 0.975 & z1 < .
(1,396 real changes made)
```

```
. replace lrej1 = z1 < -0.975
(1,041 real changes made)
```

```
. replace rrej2 = z2 > 0.975
(0 real changes made)
```

```
. replace lrej2 = z2 < -0.975
(919 real changes made)
```

```
.
. *0.175
. *
. sum rrej1-lrej2
```

Variable	Obs	Mean	Std. Dev.	Min	Max
rrej1	10,000	.1604	.366995	0	1
lrej1	10,000	.1348	.3415269	0	1
rrej2	10,000	.336	.4723625	0	1
lrej2	10,000	.0924	.2896043	0	1

```
.
. bitest lrej1 = 0.175
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej1	10000	1348	1750	0.17500	0.13480

```
Pr(k >= 1348) = 1.000000 (one-sided test)
Pr(k <= 1348) = 0.000000 (one-sided test)
Pr(k <= 1348 or k >= 2178) = 0.000000 (two-sided test)
```

```
. bitest rrej1 = 0.175
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej1	10000	1604	1750	0.17500	0.16040
Pr(k >= 1604) = 0.999951 (one-sided test)					
Pr(k <= 1604) = 0.000054 (one-sided test)					
Pr(k <= 1604 or k >= 1899) = 0.000109 (two-sided test)					

```
. bitest lrej2 = 0.175
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej2	10000	924	1750	0.17500	0.09240
Pr(k >= 924) = 1.000000 (one-sided test)					
Pr(k <= 924) = 0.000000 (one-sided test)					
Pr(k <= 924 or k >= 2696) = 0.000000 (two-sided test)					

```
. bitest rrej2 = 0.175
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej2	10000	3360	1750	0.17500	0.33600
Pr(k >= 3360) = 0.000000 (one-sided test)					
Pr(k <= 3360) = 1.000000 (one-sided test)					
Pr(k <= 473 or k >= 3360) = 0.000000 (two-sided test)					

```
.
.
. replace rrej1 = z1 > 0.025 & z1 < .
(3,504 real changes made)
```

```
. replace lrej1 = z1 < -0.025
(3,544 real changes made)
```

```
. replace rrej2 = z2 > 0.025
(3,693 real changes made)
```

```
. replace lrej2 = z2 < -0.025
(2,023 real changes made)
```

```
.
. *0.5
. *
. sum rrej1-lrej2
```

Variable	Obs	Mean	Std. Dev.	Min	Max
rrej1	10,000	.5108	.4999083	0	1
lrej1	10,000	.4892	.4999083	0	1
rrej2	10,000	.7053	.4559306	0	1
lrej2	10,000	.2947	.4559306	0	1

```
.
. bitest lrej1 = 0.50
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej1	10000	4892	5000	0.50000	0.48920
Pr(k >= 4892) = 0.984999 (one-sided test)					
Pr(k <= 4892) = 0.015775 (one-sided test)					
Pr(k <= 4892 or k >= 5108) = 0.031550 (two-sided test)					

```
. bitest rrej1 = 0.50
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej1	10000	5108	5000	0.50000	0.51080

Pr(k >= 5108) = 0.015775 (one-sided test)
 Pr(k <= 5108) = 0.984999 (one-sided test)
 Pr(k <= 4892 or k >= 5108) = 0.031550 (two-sided test)

```
. bitest lrej2 = 0.50
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej2	10000	2947	5000	0.50000	0.29470

Pr(k >= 2947) = 1.000000 (one-sided test)
 Pr(k <= 2947) = 0.000000 (one-sided test)
 Pr(k <= 2947 or k >= 7053) = 0.000000 (two-sided test)

```
. bitest rrej2 = 0.50
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej2	10000	7053	5000	0.50000	0.70530

Pr(k >= 7053) = 0.000000 (one-sided test)
 Pr(k <= 7053) = 1.000000 (one-sided test)
 Pr(k <= 2947 or k >= 7053) = 0.000000 (two-sided test)

```
.
```

```
. clear
```

```
.
```

```
. *
```

```
. *
```

```
. simulate mu1 = r(mu1) mu2 = r(mu2) sd1 = r(sd1) sd2 = r(sd2), reps(10000) no  
> dots: clt 64
```

```

  command: clt 64
    mu1:   r(mu1)
    mu2:   r(mu2)
    sd1:   r(sd1)
    sd2:   r(sd2)

```

```
.
```

```
. * save this data
```

```
.
```

```
. save clt64, replace  
file clt64.dta saved
```

```
.
```

```
. * Sample using 64 observations
```

```
.
```

```
. use clt64
```

```
(simulate: clt)
```

```
. gen z1 = (mu1-0.35) / (sd1/64^(1/2))
```

```
. gen z2 = (mu2-0.97) / (sd2/64^(1/2))
(1,419 missing values generated)
```

```
. *
. * construct rejection level 5 percent rate; left and right tails
. *
. gen rrej1 = z1 > 1.96 & z1 < .
. gen lrej1 = z1 < -1.96
. gen rrej2 = z2 > 1.96
. gen lrej2 = z2 < -1.96
. *
. * Should be 0.05, 1 binary 2 binary
. *
. sum rrej1-lrej2 // 64 observations
```

Variable	Obs	Mean	Std. Dev.	Min	Max
rrej1	10,000	.0183	.1340407	0	1
lrej1	10,000	.0332	.1791674	0	1
rrej2	10,000	.1419	.348965	0	1
lrej2	10,000	.0023	.0479055	0	1

```
. /*
> 99% (-1%) 2.33
> 95% (-5%) 1.65
> 90% (-10%) 1.29
> 75% (-25%) 0.58
> */
```

```
. bitest lrej1 = 0.025
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej1	10000	332	250	0.02500	0.03320

```
Pr(k >= 332) = 0.000000 (one-sided test)
Pr(k <= 332) = 1.000000 (one-sided test)
Pr(k <= 175 or k >= 332) = 0.000001 (two-sided test)
```

```
. bitest rrej1 = 0.025
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej1	10000	183	250	0.02500	0.01830

```
Pr(k >= 183) = 0.999997 (one-sided test)
Pr(k <= 183) = 0.000004 (one-sided test)
Pr(k <= 183 or k >= 323) = 0.000008 (two-sided test)
```

```
. bitest lrej2 = 0.025
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej2	10000	23	250	0.02500	0.00230

```
Pr(k >= 23) = 1.000000 (one-sided test)
Pr(k <= 23) = 0.000000 (one-sided test)
Pr(k <= 23 or k >= 591) = 0.000000 (two-sided test)
```

```
. bitest rrej2 = 0.025
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej2	10000	1419	250	0.02500	0.14190

```
Pr(k >= 1419) = 0.000000 (one-sided test)
Pr(k <= 1419) = 1.000000 (one-sided test)
Pr(k >= 1419) = 0.000000 (two-sided test)
```

```
note: lower tail of two-sided p-value is empty
```

```
.
.
. replace rrej1 = z1 > 0.975 & z1 < .
(1,224 real changes made)
```

```
. replace lrej1 = z1 < -0.975
(1,249 real changes made)
```

```
. replace rrej2 = z2 > 0.975
(0 real changes made)
```

```
. replace lrej2 = z2 < -0.975
(1,206 real changes made)
```

```
.
. *
. * Should be around 0.175
. *
. sum rrej1-lrej2
```

Variable	Obs	Mean	Std. Dev.	Min	Max
rrej1	10,000	.1407	.3477292	0	1
lrej1	10,000	.1581	.364853	0	1
rrej2	10,000	.1419	.348965	0	1
lrej2	10,000	.1229	.3283388	0	1

```
.
. bitest lrej1 = 0.175
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej1	10000	1581	1750	0.17500	0.15810

```
Pr(k >= 1581) = 0.999997 (one-sided test)
Pr(k <= 1581) = 0.000004 (one-sided test)
Pr(k <= 1581 or k >= 1923) = 0.000007 (two-sided test)
```

```
. bitest rrej1 = 0.175
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej1	10000	1407	1750	0.17500	0.14070

```
Pr(k >= 1407) = 1.000000 (one-sided test)
Pr(k <= 1407) = 0.000000 (one-sided test)
Pr(k <= 1407 or k >= 2112) = 0.000000 (two-sided test)
```

```
. bitest lrej2 = 0.175
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej2	10000	1229	1750	0.17500	0.12290
Pr(k >= 1229)		= 1.000000	(one-sided test)		
Pr(k <= 1229)		= 0.000000	(one-sided test)		
Pr(k <= 1229 or k >= 2315)		= 0.000000	(two-sided test)		

```
. bitest rrej2 = 0.175
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej2	10000	1419	1750	0.17500	0.14190
Pr(k >= 1419)		= 1.000000	(one-sided test)		
Pr(k <= 1419)		= 0.000000	(one-sided test)		
Pr(k <= 1419 or k >= 2098)		= 0.000000	(two-sided test)		

```
.
.
. replace rrej1 = z1 > 0.025 & z1 < .
(3,408 real changes made)
```

```
. replace lrej1 = z1 < -0.025
(3,604 real changes made)
```

```
. replace rrej2 = z2 > 0.025
(2,813 real changes made)
```

```
. replace lrej2 = z2 < -0.025
(4,539 real changes made)
```

```
.
. *
. * Should be 0.50
. *
. sum rrej1-lrej2
```

Variable	Obs	Mean	Std. Dev.	Min	Max
rrej1	10,000	.4815	.4996826	0	1
lrej1	10,000	.5185	.4996826	0	1
rrej2	10,000	.4232	.4940913	0	1
lrej2	10,000	.5768	.4940913	0	1

```
. bitest lrej1 = 0.50
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej1	10000	5185	5000	0.50000	0.51850
Pr(k >= 5185)		= 0.000112	(one-sided test)		
Pr(k <= 5185)		= 0.999897	(one-sided test)		
Pr(k <= 4815 or k >= 5185)		= 0.000224	(two-sided test)		

```
. bitest rrej1 = 0.50
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej1	10000	4815	5000	0.50000	0.48150
Pr(k >= 4815)		= 0.999897	(one-sided test)		
Pr(k <= 4815)		= 0.000112	(one-sided test)		
Pr(k <= 4815 or k >= 5185)		= 0.000224	(two-sided test)		


```
. bitest lrej2 = 0.50
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej2	10000	5768	5000	0.50000	0.57680

```
Pr(k >= 5768) = 0.000000 (one-sided test)
Pr(k <= 5768) = 1.000000 (one-sided test)
Pr(k <= 4232 or k >= 5768) = 0.000000 (two-sided test)
```

```
. bitest rrej2 = 0.50
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej2	10000	4232	5000	0.50000	0.42320

```
Pr(k >= 4232) = 1.000000 (one-sided test)
Pr(k <= 4232) = 0.000000 (one-sided test)
Pr(k <= 4232 or k >= 5768) = 0.000000 (two-sided test)
```

```
.
.
. clear

. *
. * run simulation for 100 observations
. *
. simulate mu1 = r(mu1) mu2 = r(mu2) sd1 = r(sd1) sd2 = r(sd2), reps(10000) no
> dots: clt 100
```

```
command: clt 100
mu1: r(mu1)
mu2: r(mu2)
sd1: r(sd1)
sd2: r(sd2)
```

```
. *
. * save this data
. *
. save clt100, replace
file clt100.dta saved
```

```
. *
. * Sample using 100 observations
. *
. use clt100
(simulate: clt)

. gen z1 = (mu1-0.35) / (sd1/100^(1/2))

. gen z2 = (mu2-0.97) / (sd2/100^(1/2))
(469 missing values generated)
```

```
.
. *
. * construct rejection level 5 percent rate; left and right tails
. *
. gen rrej1 = z1 > 1.96 & z1 < .
```

```
. gen lrej1 = z1 < -1.96
. gen rrej2 = z2 > 1.96
. gen lrej2 = z2 < -1.96

. *
. * Should be 0.05, 1 binary 2 binary
. *
. sum rrej1-lrej2 // 100 observations
```

Variable	Obs	Mean	Std. Dev.	Min	Max
rrej1	10,000	.0237	.1521204	0	1
lrej1	10,000	.0365	.18754	0	1
rrej2	10,000	.1968	.3975998	0	1
lrej2	10,000	.0033	.0573536	0	1

```
. /*
> 99% (-1%) 2.33
> 95% (-5%) 1.65
> 90% (-10%) 1.29
> 75% (-25%) 0.58
> */
```

```
. bitest lrej1 = 0.025
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej1	10000	365	250	0.02500	0.03650

```
Pr(k >= 365) = 0.000000 (one-sided test)
Pr(k <= 365) = 1.000000 (one-sided test)
Pr(k <= 149 or k >= 365) = 0.000000 (two-sided test)
```

```
. bitest rrej1 = 0.025
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej1	10000	237	250	0.02500	0.02370

```
Pr(k >= 237) = 0.805713 (one-sided test)
Pr(k <= 237) = 0.212713 (one-sided test)
Pr(k <= 237 or k >= 263) = 0.423316 (two-sided test)
```

```
. bitest lrej2 = 0.025
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej2	10000	33	250	0.02500	0.00330

```
Pr(k >= 33) = 1.000000 (one-sided test)
Pr(k <= 33) = 0.000000 (one-sided test)
Pr(k <= 33 or k >= 566) = 0.000000 (two-sided test)
```

```
. bitest rrej2 = 0.025
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej2	10000	1968	250	0.02500	0.19680

```
Pr(k >= 1968) = 0.000000 (one-sided test)
Pr(k <= 1968) = 1.000000 (one-sided test)
Pr(k >= 1968) = 0.000000 (two-sided test)
```

note: lower tail of two-sided p-value is empty

```

.
.
. replace rrej1 = z1 > 0.975 & z1 < .
(1,467 real changes made)

. replace lrej1 = z1 < -0.975
(1,309 real changes made)

. replace rrej2 = z2 > 0.975
(0 real changes made)

. replace lrej2 = z2 < -0.975
(760 real changes made)

.
. *
. * Should be around 0.175
. *
. sum rrej1-lrej2

```

Variable	Obs	Mean	Std. Dev.	Min	Max
rrej1	10,000	.1704	.3760026	0	1
lrej1	10,000	.1674	.3733513	0	1
rrej2	10,000	.1968	.3975998	0	1
lrej2	10,000	.0793	.2702199	0	1

```

. bitest lrej1 = 0.175

```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej1	10000	1674	1750	0.17500	0.16740

```

Pr(k >= 1674) = 0.978424 (one-sided test)
Pr(k <= 1674) = 0.022988 (one-sided test)
Pr(k <= 1674 or k >= 1827) = 0.045479 (two-sided test)

```

```

. bitest rrej1 = 0.175

```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej1	10000	1704	1750	0.17500	0.17040

```

Pr(k >= 1704) = 0.889758 (one-sided test)
Pr(k <= 1704) = 0.115315 (one-sided test)
Pr(k <= 1704 or k >= 1796) = 0.231111 (two-sided test)

```

```

. bitest lrej2 = 0.175

```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej2	10000	793	1750	0.17500	0.07930

```

Pr(k >= 793) = 1.000000 (one-sided test)
Pr(k <= 793) = 0.000000 (one-sided test)
Pr(k <= 793 or k >= 2875) = 0.000000 (two-sided test)

```

```

. bitest rrej2 = 0.175

```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej2	10000	1968	1750	0.17500	0.19680

```

Pr(k >= 1968) = 0.000000 (one-sided test)
Pr(k <= 1968) = 1.000000 (one-sided test)
Pr(k <= 1538 or k >= 1968) = 0.000000 (two-sided test)

```

```
.
.
. replace rrej1 = z1 > 0.025 & z1 < .
(2,839 real changes made)
```

```
. replace lrej1 = z1 < -0.025
(2,949 real changes made)
```

```
. replace rrej2 = z2 > 0.025
(2,216 real changes made)
```

```
. replace lrej2 = z2 < -0.025
(2,688 real changes made)
```

```
.
. *
. * Should be 0.50
. *
. sum rrej1-lrej2
```

Variable	Obs	Mean	Std. Dev.	Min	Max
rrej1	10,000	.4543	.497932	0	1
lrej1	10,000	.4623	.4986016	0	1
rrej2	10,000	.4184	.4933212	0	1
lrej2	10,000	.3481	.4763917	0	1

```
. bitest lrej1 = 0.50
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej1	10000	4623	5000	0.50000	0.46230

```
Pr(k >= 4623) = 1.000000 (one-sided test)
Pr(k <= 4623) = 0.000000 (one-sided test)
Pr(k <= 4623 or k >= 5377) = 0.000000 (two-sided test)
```

```
. bitest rrej1 = 0.50
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej1	10000	4543	5000	0.50000	0.45430

```
Pr(k >= 4543) = 1.000000 (one-sided test)
Pr(k <= 4543) = 0.000000 (one-sided test)
Pr(k <= 4543 or k >= 5457) = 0.000000 (two-sided test)
```

```
. bitest lrej2 = 0.50
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej2	10000	3481	5000	0.50000	0.34810

```
Pr(k >= 3481) = 1.000000 (one-sided test)
Pr(k <= 3481) = 0.000000 (one-sided test)
Pr(k <= 3481 or k >= 6519) = 0.000000 (two-sided test)
```

```
. bitest rrej2 = 0.50
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej2	10000	4184	5000	0.50000	0.41840

```
Pr(k >= 4184) = 1.000000 (one-sided test)
Pr(k <= 4184) = 0.000000 (one-sided test)
Pr(k <= 4184 or k >= 5816) = 0.000000 (two-sided test)
```

```

.
.
. clear

.
. *
. * run simulation for 225 observations
. *
. simulate mu1 = r(mu1) mu2 = r(mu2) sd1 = r(sd1) sd2 = r(sd2), reps(10000) no
> dots: clt 225

      command:  clt 225
             mu1:  r(mu1)
             mu2:  r(mu2)
             sd1:  r(sd1)
             sd2:  r(sd2)

.
. *
. * save this data
. *
. save clt225, replace
file clt225.dta saved

.
. *
. * Sample using 225 observations
. *
. use clt225
(simulate: clt)

. gen z1 = (mu1-0.35)      / (sd1/225^(1/2))

. gen z2 = (mu2-0.97)      / (sd2/225^(1/2))
(11 missing values generated)

.
. *
. * construct rejection level 5 percent rate; left and right tails
. *
. gen rrej1 = z1 >  1.96 & z1 < .

. gen lrej1 = z1 < -1.96

. gen rrej2 = z2 >  1.96

. gen lrej2 = z2 < -1.96

.
. *
. * Should be 0.05, 1 binary 2 binary
. *
. sum rrej1-lrej2 // 225 observations

      Variable |      Obs      Mean   Std. Dev.      Min      Max
-----+-----+-----+-----+-----+-----+-----
      rrej1 |    10,000    .0203    .1410316         0         1
      lrej1 |    10,000    .0312    .1738666         0         1
      rrej2 |    10,000    .0931    .2905871         0         1
      lrej2 |    10,000    .0083     .09073         0         1

. /*
> 99% (-1%)  2.33
> 95% (-5%)  1.65
> 90% (-10%) 1.29
> 75% (-25%) 0.58
> */

```

```

. bitest lrej1 = 0.025

  Variable |      N   Observed k   Expected k   Assumed p   Observed p
-----+-----
    lrej1 |   10000         312         250       0.02500       0.03120

Pr(k >= 312)          = 0.000071 (one-sided test)
Pr(k <= 312)          = 0.999945 (one-sided test)
Pr(k <= 191 or k >= 312) = 0.000120 (two-sided test)

. bitest rrej1 = 0.025

  Variable |      N   Observed k   Expected k   Assumed p   Observed p
-----+-----
    rrej1 |   10000         203         250       0.02500       0.02030

Pr(k >= 203)          = 0.999138 (one-sided test)
Pr(k <= 203)          = 0.001086 (one-sided test)
Pr(k <= 203 or k >= 300) = 0.002100 (two-sided test)

. bitest lrej2 = 0.025

  Variable |      N   Observed k   Expected k   Assumed p   Observed p
-----+-----
    lrej2 |   10000          83         250       0.02500       0.00830

Pr(k >= 83)           = 1.000000 (one-sided test)
Pr(k <= 83)           = 0.000000 (one-sided test)
Pr(k <= 83 or k >= 466) = 0.000000 (two-sided test)

. bitest rrej2 = 0.025

  Variable |      N   Observed k   Expected k   Assumed p   Observed p
-----+-----
    rrej2 |   10000         931         250       0.02500       0.09310

Pr(k >= 931) = 0.000000 (one-sided test)
Pr(k <= 931) = 1.000000 (one-sided test)
Pr(k >= 931) = 0.000000 (two-sided test)

note: lower tail of two-sided p-value is empty

.
.
. replace rrej1 = z1 > 0.975 & z1 < .
(1,509 real changes made)

. replace lrej1 = z1 < -0.975
(1,282 real changes made)

. replace rrej2 = z2 > 0.975
(1,019 real changes made)

. replace lrej2 = z2 < -0.975
(1,293 real changes made)

.
. *
. * Should be around 0.175

```

```
. *
. sum rrej1-lrej2
```

Variable	Obs	Mean	Std. Dev.	Min	Max
rrej1	10,000	.1712	.3767025	0	1
lrej1	10,000	.1594	.366067	0	1
rrej2	10,000	.195	.3962205	0	1
lrej2	10,000	.1376	.3444969	0	1

```
.
. bitest lrej1 = 0.175
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej1	10000	1594	1750	0.17500	0.15940

```
Pr(k >= 1594)          = 0.999985 (one-sided test)
Pr(k <= 1594)          = 0.000017 (one-sided test)
Pr(k <= 1594 or k >= 1910) = 0.000034 (two-sided test)
```

```
. bitest rrej1 = 0.175
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej1	10000	1712	1750	0.17500	0.17120

```
Pr(k >= 1712)          = 0.844550 (one-sided test)
Pr(k <= 1712)          = 0.161854 (one-sided test)
Pr(k <= 1712 or k >= 1788) = 0.323673 (two-sided test)
```

```
. bitest lrej2 = 0.175
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej2	10000	1376	1750	0.17500	0.13760

```
Pr(k >= 1376)          = 1.000000 (one-sided test)
Pr(k <= 1376)          = 0.000000 (one-sided test)
Pr(k <= 1376 or k >= 2146) = 0.000000 (two-sided test)
```

```
. bitest rrej2 = 0.175
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej2	10000	1950	1750	0.17500	0.19500

```
Pr(k >= 1950)          = 0.000000 (one-sided test)
Pr(k <= 1950)          = 1.000000 (one-sided test)
Pr(k <= 1555 or k >= 1950) = 0.000000 (two-sided test)
```

```
.
.
. replace rrej1 = z1 > 0.025 & z1 < .
(3,382 real changes made)
```

```
. replace lrej1 = z1 < -0.025
(3,312 real changes made)
```

```
. replace rrej2 = z2 > 0.025
(2,940 real changes made)
```

```
. replace lrej2 = z2 < -0.025
(3,734 real changes made)
```

```
. *
. * Should be 0.50
. *
. sum rrej1-lrej2
```

Variable	Obs	Mean	Std. Dev.	Min	Max
rrej1	10,000	.5094	.4999366	0	1
lrej1	10,000	.4906	.4999366	0	1
rrej2	10,000	.489	.499904	0	1
lrej2	10,000	.511	.499904	0	1

```
. bitest lrej1 = 0.50
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej1	10000	4906	5000	0.50000	0.49060

Pr(k >= 4906) = 0.970624 (one-sided test)
 Pr(k <= 4906) = 0.030739 (one-sided test)
 Pr(k <= 4906 or k >= 5094) = 0.061478 (two-sided test)

```
. bitest rrej1 = 0.50
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej1	10000	5094	5000	0.50000	0.50940

Pr(k >= 5094) = 0.030739 (one-sided test)
 Pr(k <= 5094) = 0.970624 (one-sided test)
 Pr(k <= 4906 or k >= 5094) = 0.061478 (two-sided test)

```
. bitest lrej2 = 0.50
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej2	10000	5110	5000	0.50000	0.51100

Pr(k >= 5110) = 0.014260 (one-sided test)
 Pr(k <= 5110) = 0.986450 (one-sided test)
 Pr(k <= 4890 or k >= 5110) = 0.028519 (two-sided test)

```
. bitest rrej2 = 0.50
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej2	10000	4890	5000	0.50000	0.48900

Pr(k >= 4890) = 0.986450 (one-sided test)
 Pr(k <= 4890) = 0.014260 (one-sided test)
 Pr(k <= 4890 or k >= 5110) = 0.028519 (two-sided test)

```
.
.
. clear
```



```
.
. *
. * run simulation for 2500 observations
. *
. simulate mu1 = r(mu1) mu2 = r(mu2) sd1 = r(sd1) sd2 = r(sd2), reps(10000) no
> dots: clt 2500
```

```
command: clt 2500
mu1: r(mu1)
mu2: r(mu2)
sd1: r(sd1)
sd2: r(sd2)
```

```
. *
. * save this data
. *
. save clt2500, replace
file clt2500.dta saved
```

```
. *
. * Sample using 2500 observations
. *
. use clt2500
(simulate: clt)
```

```
. gen z1 = (mu1-0.35) / (sd1/2500^(1/2))
. gen z2 = (mu2-0.97) / (sd2/2500^(1/2))
```

```
. *
. * construct rejection level 5 percent rate; left and right tails
. *
. gen rrej1 = z1 > 1.96 & z1 < .
. gen lrej1 = z1 < -1.96
. gen rrej2 = z2 > 1.96
. gen lrej2 = z2 < -1.96

. *
. * Should be 0.05, 1 binary 2 binary
. *
. sum rrej1-lrej2 // 2500 observations
```

Variable	Obs	Mean	Std. Dev.	Min	Max
rrej1	10,000	.0242	.1536773	0	1
lrej1	10,000	.0244	.1542952	0	1
rrej2	10,000	.0294	.1689334	0	1
lrej2	10,000	.0176	.1314989	0	1

```
. /*
> 99% (-1%) 2.33
> 95% (-5%) 1.65
> 90% (-10%) 1.29
> 75% (-25%) 0.58
> */
```

```
. bitest lrej1 = 0.025
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej1	10000	244	250	0.02500	0.02440
Pr(k >= 244) = 0.658333 (one-sided test)					
Pr(k <= 244) = 0.365661 (one-sided test)					
Pr(k <= 244 or k >= 256) = 0.724658 (two-sided test)					

```
. bitest rrej1 = 0.025
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej1	10000	242	250	0.02500	0.02420
Pr(k >= 242) = 0.704462 (one-sided test)					
Pr(k <= 242) = 0.318266 (one-sided test)					
Pr(k <= 242 or k >= 258) = 0.630982 (two-sided test)					

```
. bitest lrej2 = 0.025
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej2	10000	176	250	0.02500	0.01760
Pr(k >= 176) = 1.000000 (one-sided test)					
Pr(k <= 176) = 0.000000 (one-sided test)					
Pr(k <= 176 or k >= 331) = 0.000001 (two-sided test)					

```
. bitest rrej2 = 0.025
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej2	10000	294	250	0.02500	0.02940
Pr(k >= 294) = 0.003239 (one-sided test)					
Pr(k <= 294) = 0.997310 (one-sided test)					
Pr(k <= 207 or k >= 294) = 0.005851 (two-sided test)					

```
.
.
. replace rrej1 = z1 > 0.975 & z1 < .
(1,366 real changes made)
```

```
. replace lrej1 = z1 < -0.975
(1,413 real changes made)
```

```
. replace rrej2 = z2 > 0.975
(1,624 real changes made)
```

```
. replace lrej2 = z2 < -0.975
(1,452 real changes made)
```

```
.
. *
. * Should be around 0.175
. *
. sum rrej1-lrej2
```

Variable	Obs	Mean	Std. Dev.	Min	Max
rrej1	10,000	.1608	.3673647	0	1
lrej1	10,000	.1657	.3718297	0	1
rrej2	10,000	.1918	.3937363	0	1
lrej2	10,000	.1628	.3692016	0	1

```
. bitest lrej1 = 0.175
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej1	10000	1657	1750	0.17500	0.16570
Pr(k >= 1657) = 0.993348 (one-sided test)					
Pr(k <= 1657) = 0.007166 (one-sided test)					
Pr(k <= 1657 or k >= 1844) = 0.014374 (two-sided test)					

```
. bitest rrej1 = 0.175
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej1	10000	1608	1750	0.17500	0.16080
Pr(k >= 1608) = 0.999924 (one-sided test)					
Pr(k <= 1608) = 0.000084 (one-sided test)					
Pr(k <= 1608 or k >= 1895) = 0.000167 (two-sided test)					

```
. bitest lrej2 = 0.175
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej2	10000	1628	1750	0.17500	0.16280
Pr(k >= 1628) = 0.999426 (one-sided test)					
Pr(k <= 1628) = 0.000630 (one-sided test)					
Pr(k <= 1628 or k >= 1874) = 0.001262 (two-sided test)					

```
. bitest rrej2 = 0.175
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej2	10000	1918	1750	0.17500	0.19180
Pr(k >= 1918) = 0.000007 (one-sided test)					
Pr(k <= 1918) = 0.999994 (one-sided test)					
Pr(k <= 1585 or k >= 1918) = 0.000012 (two-sided test)					

```
.
.
. replace rrej1 = z1 > 0.025 & z1 < .
(3,304 real changes made)
```

```
. replace lrej1 = z1 < -0.025
(3,254 real changes made)
```

```
. replace rrej2 = z2 > 0.025
(2,907 real changes made)
```

```
. replace lrej2 = z2 < -0.025
(3,057 real changes made)
```

```
.
. *
. * Should be 0.50
. *
. sum rrej1-lrej2
```

Variable	Obs	Mean	Std. Dev.	Min	Max
rrej1	10,000	.4912	.4999476	0	1
lrej1	10,000	.4911	.4999458	0	1
rrej2	10,000	.4825	.4997186	0	1
lrej2	10,000	.4685	.4990317	0	1

```

. bitest lrej1 = 0.50

```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej1	10000	4911	5000	0.50000	0.49110

```

Pr(k >= 4911) = 0.963276 (one-sided test)
Pr(k <= 4911) = 0.038361 (one-sided test)
Pr(k <= 4911 or k >= 5089) = 0.076722 (two-sided test)

```

```

. bitest rrej1 = 0.50

```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej1	10000	4912	5000	0.50000	0.49120

```

Pr(k >= 4912) = 0.961639 (one-sided test)
Pr(k <= 4912) = 0.040057 (one-sided test)
Pr(k <= 4912 or k >= 5088) = 0.080113 (two-sided test)

```

```

. bitest lrej2 = 0.50

```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej2	10000	4685	5000	0.50000	0.46850

```

Pr(k >= 4685) = 1.000000 (one-sided test)
Pr(k <= 4685) = 0.000000 (one-sided test)
Pr(k <= 4685 or k >= 5315) = 0.000000 (two-sided test)

```

```

. bitest rrej2 = 0.50

```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej2	10000	4825	5000	0.50000	0.48250

```

Pr(k >= 4825) = 0.999776 (one-sided test)
Pr(k <= 4825) = 0.000241 (one-sided test)
Pr(k <= 4825 or k >= 5175) = 0.000482 (two-sided test)

```

```

.
. clear

.
. *
. * run simulation for 12100 observations
. *
. simulate mu1 = r(mu1) mu2 = r(mu2) sd1 = r(sd1) sd2 = r(sd2), reps(10000) no
> dots: clt 12100

      command:  clt 12100
              mu1:  r(mu1)
              mu2:  r(mu2)
              sd1:  r(sd1)
              sd2:  r(sd2)

.
. *
. * save this data

```

```

. *
. save clt12100, replace
file clt12100.dta saved

. *
. *
. * Sample using 12100 observations
. *
. use clt12100
(simulate: clt)

. gen z1 = (mu1-0.35)          /(sd1/12100^(1/2))
. gen z2 = (mu2-0.97)          /(sd2/12100^(1/2))

. *
. *
. * construct rejection level 5 percent rate; left and right tails
. *
. gen rrej1 = z1 > 1.96 & z1 < .
. gen lrej1 = z1 < -1.96
. gen rrej2 = z2 > 1.96
. gen lrej2 = z2 < -1.96

. *
. * Should be 0.05
. *
. sum rrej1-lrej2 // 12100 observations

```

Variable	Obs	Mean	Std. Dev.	Min	Max
rrej1	10,000	.0235	.1514927	0	1
lrej1	10,000	.0239	.1527453	0	1
rrej2	10,000	.0328	.1781217	0	1
lrej2	10,000	.0213	.1443897	0	1

```

.
. bitest lrej1 = 0.025

```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej1	10000	239	250	0.02500	0.02390

```

Pr(k >= 239)          = 0.767907 (one-sided test)
Pr(k <= 239)          = 0.252391 (one-sided test)
Pr(k <= 239 or k >= 261) = 0.501244 (two-sided test)

```

```

. bitest rrej1 = 0.025

```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej1	10000	235	250	0.02500	0.02350

```

Pr(k >= 235)          = 0.839596 (one-sided test)
Pr(k <= 235)          = 0.176844 (one-sided test)
Pr(k <= 235 or k >= 265) = 0.352966 (two-sided test)

```

```
. bitest lrej2 = 0.025
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej2	10000	213	250	0.02500	0.02130
Pr(k >= 213)		= 0.992936	(one-sided test)		
Pr(k <= 213)		= 0.008503	(one-sided test)		
Pr(k <= 213 or k >= 288)		= 0.017722	(two-sided test)		

```
. bitest rrej2 = 0.025
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej2	10000	328	250	0.02500	0.03280
Pr(k >= 328)		= 0.000001	(one-sided test)		
Pr(k <= 328)		= 0.999999	(one-sided test)		
Pr(k <= 178 or k >= 328)		= 0.000002	(two-sided test)		

```
.
.
. replace rrej1 = z1 > 0.975 & z1 < .
(1,404 real changes made)
```

```
. replace lrej1 = z1 < -0.975
(1,330 real changes made)
```

```
. replace rrej2 = z2 > 0.975
(1,492 real changes made)
```

```
. replace lrej2 = z2 < -0.975
(1,423 real changes made)
```

```
.
. *
. * Should be around 0.175
. *
. sum rrej1-lrej2
```

Variable	Obs	Mean	Std. Dev.	Min	Max
rrej1	10,000	.1639	.3702033	0	1
lrej1	10,000	.1569	.3637246	0	1
rrej2	10,000	.182	.3858638	0	1
lrej2	10,000	.1636	.3699307	0	1

```
. bitest lrej1 = 0.175
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej1	10000	1569	1750	0.17500	0.15690
Pr(k >= 1569)		= 0.999999	(one-sided test)		
Pr(k <= 1569)		= 0.000001	(one-sided test)		
Pr(k <= 1569 or k >= 1936)		= 0.000001	(two-sided test)		

```
. bitest rrej1 = 0.175
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej1	10000	1639	1750	0.17500	0.16390
Pr(k >= 1639)		= 0.998446	(one-sided test)		
Pr(k <= 1639)		= 0.001694	(one-sided test)		
Pr(k <= 1639 or k >= 1863)		= 0.003339	(two-sided test)		

```
. bitest lrej2 = 0.175
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej2	10000	1636	1750	0.17500	0.16360
Pr(k >= 1636)		= 0.998806	(one-sided test)		
Pr(k <= 1636)		= 0.001304	(one-sided test)		
Pr(k <= 1636 or k >= 1866)		= 0.002581	(two-sided test)		

```
. bitest rrej2 = 0.175
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej2	10000	1820	1750	0.17500	0.18200
Pr(k >= 1820)		= 0.034185	(one-sided test)		
Pr(k <= 1820)		= 0.967743	(one-sided test)		
Pr(k <= 1680 or k >= 1820)		= 0.067368	(two-sided test)		

```
.
.
. replace rrej1 = z1 > 0.025 & z1 < .
(3,308 real changes made)
```

```
. replace lrej1 = z1 < -0.025
(3,238 real changes made)
```

```
. replace rrej2 = z2 > 0.025
(3,120 real changes made)
```

```
. replace lrej2 = z2 < -0.025
(3,193 real changes made)
```

```
.
. *
. * Should be 0.50
. *
. sum rrej1-lrej2
```

Variable	Obs	Mean	Std. Dev.	Min	Max
rrej1	10,000	.4947	.4999969	0	1
lrej1	10,000	.4807	.4996524	0	1
rrej2	10,000	.494	.499989	0	1
lrej2	10,000	.4829	.4997325	0	1

```
. bitest lrej1 = 0.50
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej1	10000	4807	5000	0.50000	0.48070
Pr(k >= 4807)		= 0.999946	(one-sided test)		
Pr(k <= 4807)		= 0.000059	(one-sided test)		
Pr(k <= 4807 or k >= 5193)		= 0.000118	(two-sided test)		

```
. bitest rrej1 = 0.50
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej1	10000	4947	5000	0.50000	0.49470
Pr(k >= 4947)		= 0.857691	(one-sided test)		
Pr(k <= 4947)		= 0.146859	(one-sided test)		
Pr(k <= 4947 or k >= 5053)		= 0.293718	(two-sided test)		

```
. bitest lrej2 = 0.50
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
lrej2	10000	4829	5000	0.50000	0.48290

```
Pr(k >= 4829)          = 0.999699 (one-sided test)
Pr(k <= 4829)          = 0.000324 (one-sided test)
Pr(k <= 4829 or k >= 5171) = 0.000649 (two-sided test)
```

```
. bitest rrej2 = 0.50
```

Variable	N	Observed k	Expected k	Assumed p	Observed p
rrej2	10000	4940	5000	0.50000	0.49400

```
Pr(k >= 4940)          = 0.886861 (one-sided test)
Pr(k <= 4940)          = 0.117022 (one-sided test)
Pr(k <= 4940 or k >= 5060) = 0.234045 (two-sided test)
```

```
.
. clear
```

```
.
. *Observations
. * As we increase N, the percentage of sample means that have a z-score below
> -0.025 and above 0.025) is ~99%.
. * As we increase N, the percentage of sample means that have a z-score below
> -0.975 and above 0.975) is ~34%.
. * As we increase N, the percentage of sample means that have a z-score below
> -1.96 and above 1.96 is ~5%.
.
. *
. * close log file
. *
. log close
.   name: <unnamed>
.   log: C:\Users\saiomkark\OneDrive - The University of Chicago\AdvStats\
> PS5\Question2.log
.   log type: text
.   closed on: 5 Nov 2021, 19:32:20
-----
```