The UNIVARIATE Procedure
Variable: Gr_Liv_Area

Moments			
N	2930	**Sum Weights**	2930
Mean	1499.69044	**Sum Observations**	4394093
Std Deviation	505.508887	**Variance**	255539.235
Skewness	1.27410972	**Kurtosis**	4.13783819
Uncorrected SS	7338253701	**Corrected SS**	748474420
Coeff Variation	33.7075488	**Std Error Mean**	9.33888409

Basic Statistical Measures			
Location		**Variability**	
Mean	1499.690	**Std Deviation**	505.50889
Median	1442.000	**Variance**	255539
Mode	864.000	**Range**	5308
		Interquartile Range	617.00000

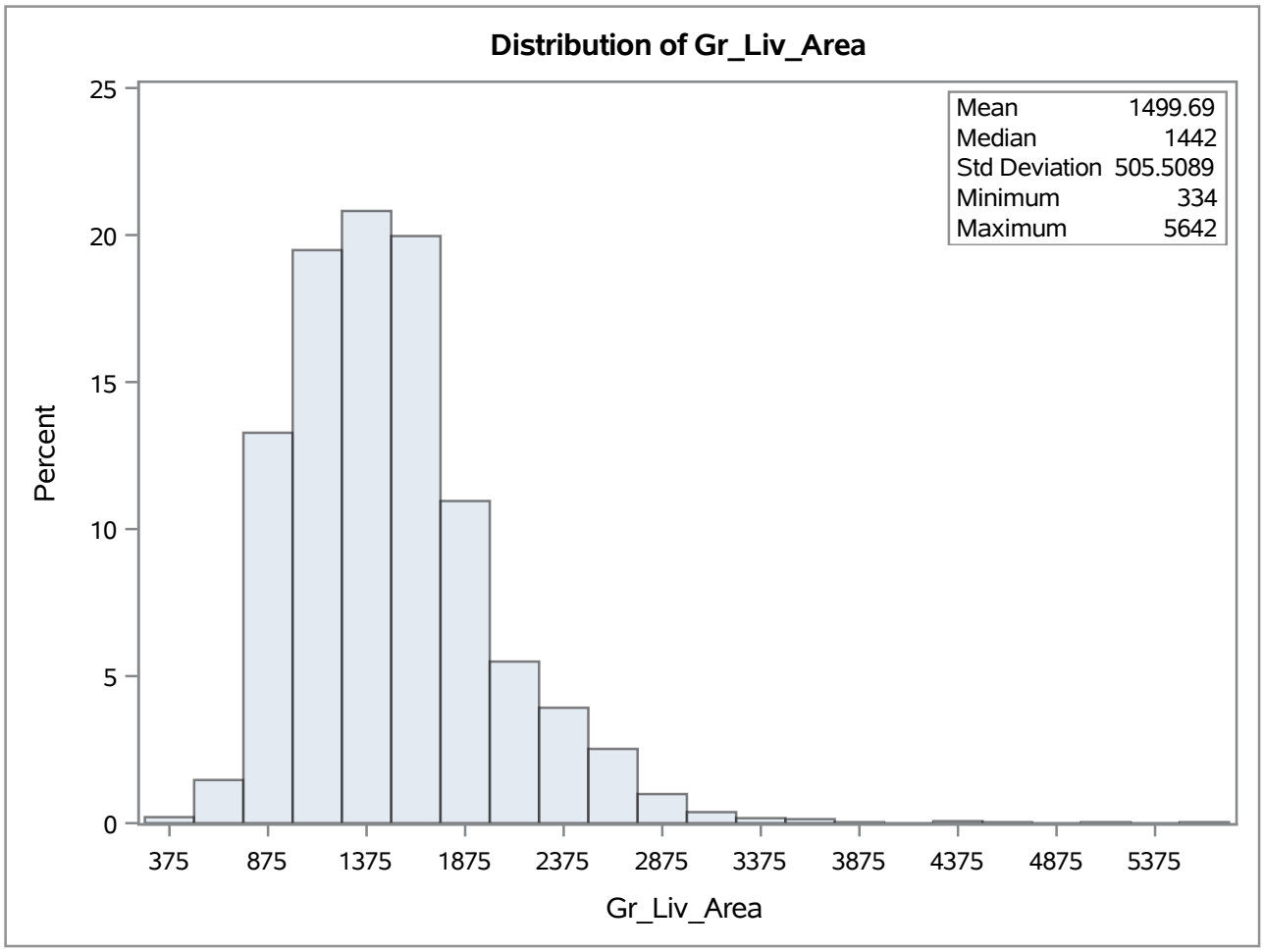
Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	160.5856	**Pr > |t|**	<.0001
Sign	**M**	1465	**Pr >= |M|**	<.0001
Signed Rank	**S**	2146958	**Pr >= |S|**	<.0001

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	5642.0
99%	2944.0
95%	2464.0
90%	2152.5
75% Q3	1743.0
50% Median	1442.0
25% Q1	1126.0
10%	923.5
5%	861.0
1%	672.0
0% Min	334.0

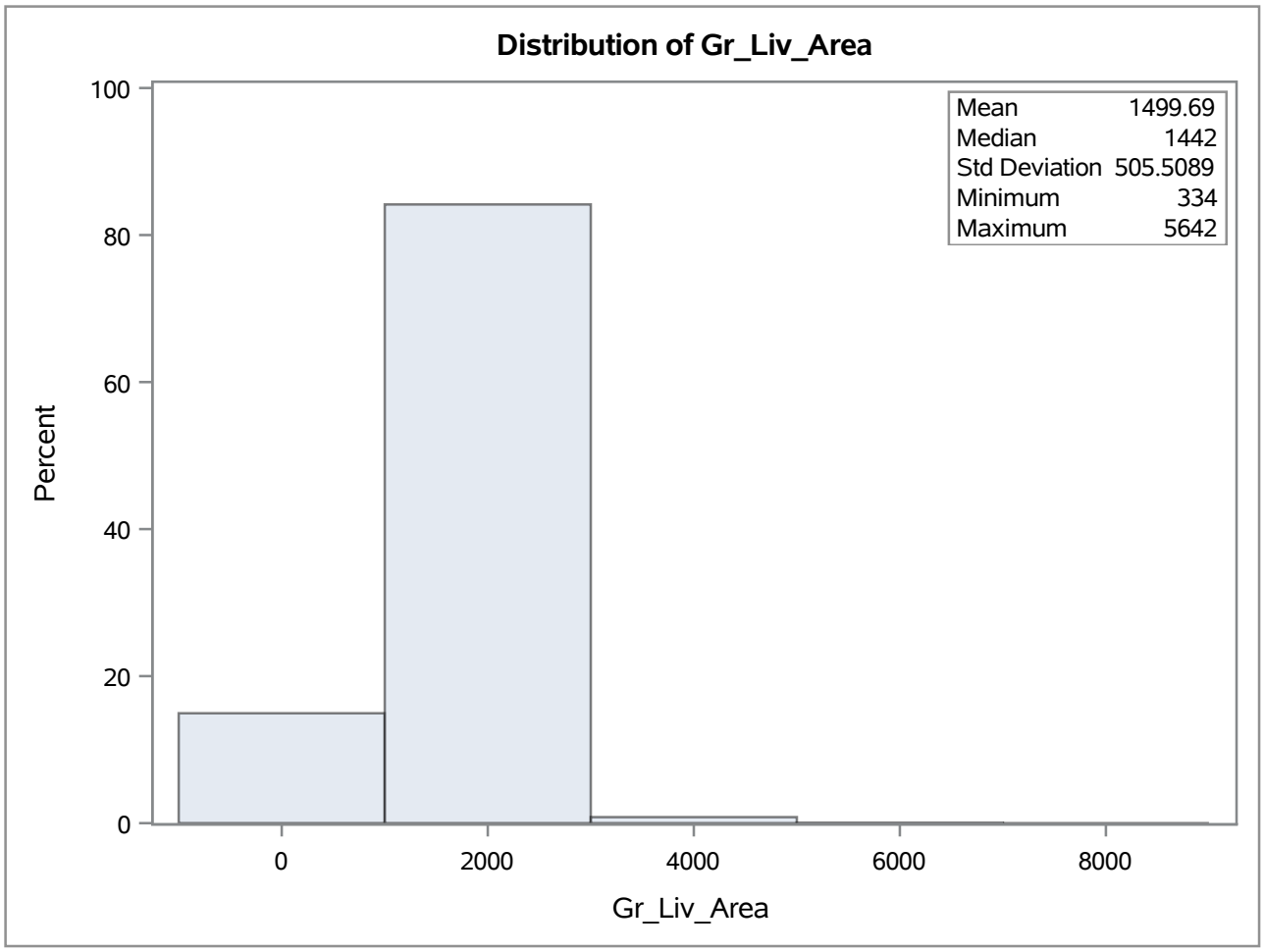
The UNIVARIATE Procedure
Variable: Gr_Liv_Area

Extreme Observations			
Lowest		**Highest**	
Value	**Obs**	**Value**	**Obs**
334	1902	4316	1768
407	1303	4476	1761
438	908	4676	2182
480	2881	5095	2181
492	2654	5642	1499

The UNIVARIATE Procedure



The UNIVARIATE Procedure



Answer
This is a Right-Skewed distribution and unimodal as we can see from the histogram.

The SURVEYSELECT Procedure

Selection Method	Simple Random Sampling

Input Data Set	AMES
Random Number Seed	39361131
Sample Size	50
Selection Probability	0.017065
Sampling Weight	58.6
Output Data Set	AMESSAMPLE

Obs	Gr_Liv_Area	SalePrice
1	1822	259000
2	1145	160000
3	1251	119000
4	1433	270000
5	992	155891
6	1968	269500
7	1472	187500
8	1630	189500
9	1850	248000
10	1844	257000
11	1319	123000
12	1152	159900
13	904	61000
14	1523	158000
15	1567	196000
16	1414	176500
17	2084	385000
18	1162	170000
19	1080	136000
20	1588	65000
21	1160	152500
22	889	137500
23	2541	349265
24	1828	314813
25	2656	492000
26	2643	380000
27	1536	176000
28	1266	185485
29	4476	745000
30	1795	147000
31	1528	235876
32	888	120000
33	3086	200500
34	1479	226000
35	894	110000
36	694	90350
37	1501	244000
38	1756	204000

Obs	Gr_Liv_Area	SalePrice
39	1178	135000
40	1083	172000
41	1578	133000
42	1375	140000
43	1008	139500
44	1207	147000
45	2486	220000
46	934	132000
47	1734	260000
48	1709	130000
49	1339	180000
50	1800	312500

Answer

Since the samples vary each time we cannot say for sure, but in this case it is a Right-Skewed distribution and follows the population distribution,also having a mean close to the population distribution.The only differences is that the range is not as wide and the outliers are not as extreme.

The UNIVARIATE Procedure
Variable: Gr_Liv_Area

Moments			
N	50	**Sum Weights**	50
Mean	1584.94	**Sum Observations**	79247
Std Deviation	658.729588	**Variance**	433924.67
Skewness	2.10576302	**Kurtosis**	6.78109374
Uncorrected SS	146864049	**Corrected SS**	21262308.8
Coeff Variation	41.5617997	**Std Error Mean**	93.1584317

Basic Statistical Measures			
Location		**Variability**	
Mean	1584.940	**Std Deviation**	658.72959
Median	1490.000	**Variance**	433925
Mode	.	**Range**	3782
		Interquartile Range	640.00000

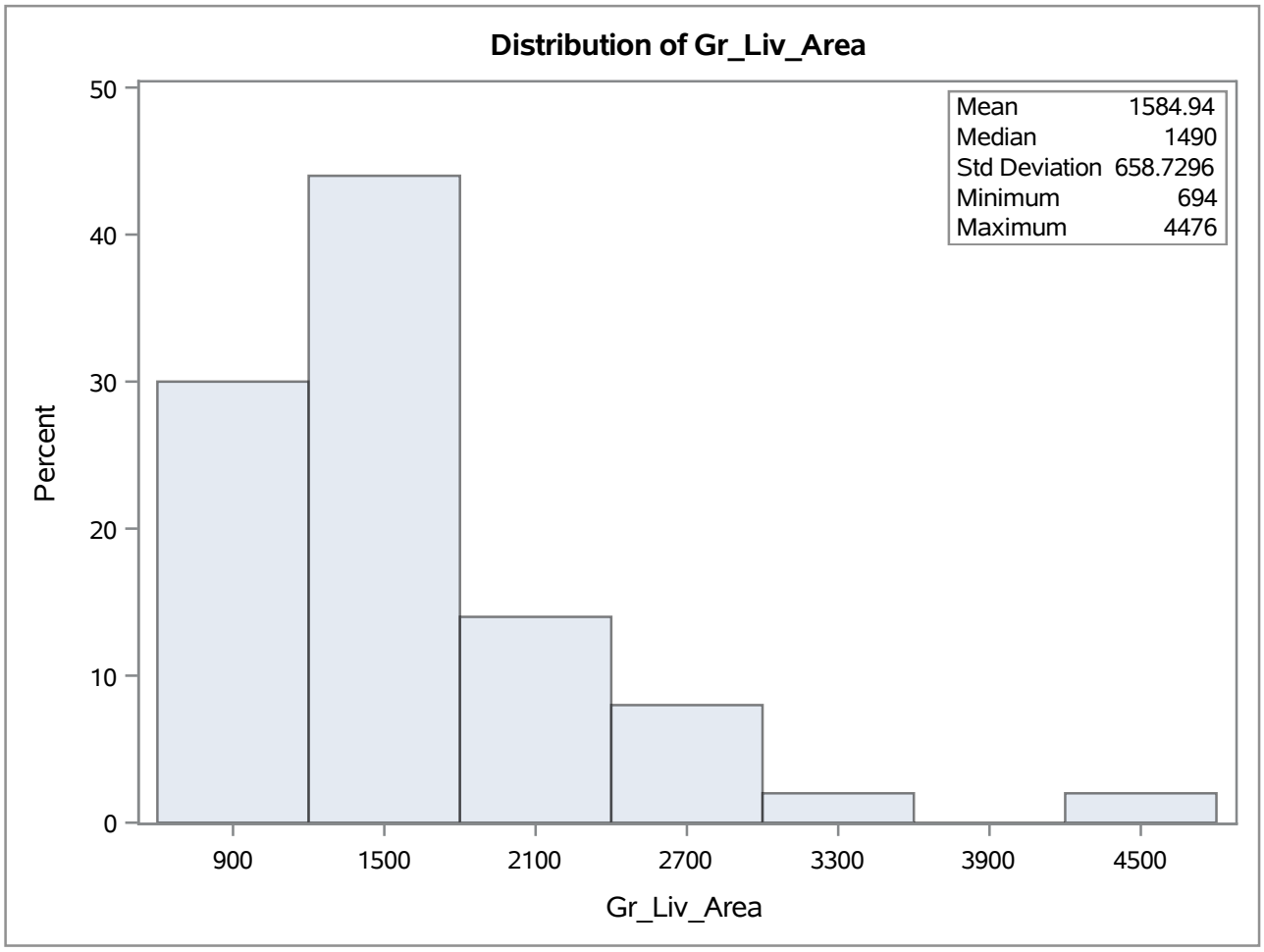
Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	17.01338	**Pr > |t|**	<.0001
Sign	**M**	25	**Pr >= |M|**	<.0001
Signed Rank	**S**	637.5	**Pr >= |S|**	<.0001

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	4476.0
99%	4476.0
95%	2656.0
90%	2513.5
75% Q3	1800.0
50% Median	1490.0
25% Q1	1160.0
10%	919.0
5%	889.0
1%	694.0
0% Min	694.0

The UNIVARIATE Procedure
Variable: Gr_Liv_Area

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
694	36	2541	23
888	32	2643	26
889	22	2656	25
894	35	3086	33
904	13	4476	29

The UNIVARIATE Procedure



The MEANS Procedure

Analysis Variable : Gr_Liv_Area
Mean
1584.94

The SURVEYSELECT Procedure

Selection Method	Simple Random Sampling

Input Data Set	AMES
Random Number Seed	39953873
Sample Size	50
Selection Probability	0.017065
Sampling Weight	58.6
Output Data Set	AMESSAMPLE2

The UNIVARIATE Procedure
Variable: Gr_Liv_Area

Moments			
N	50	**Sum Weights**	50
Mean	1489.14	**Sum Observations**	74457
Std Deviation	470.289881	**Variance**	221172.572
Skewness	0.54473268	**Kurtosis**	-0.4356561
Uncorrected SS	121714353	**Corrected SS**	10837456
Coeff Variation	31.5813074	**Std Error Mean**	66.5090327

Basic Statistical Measures			
Location		**Variability**	
Mean	1489.140	**Std Deviation**	470.28988
Median	1423.000	**Variance**	221173
Mode	.	**Range**	1767
		Interquartile Range	590.00000

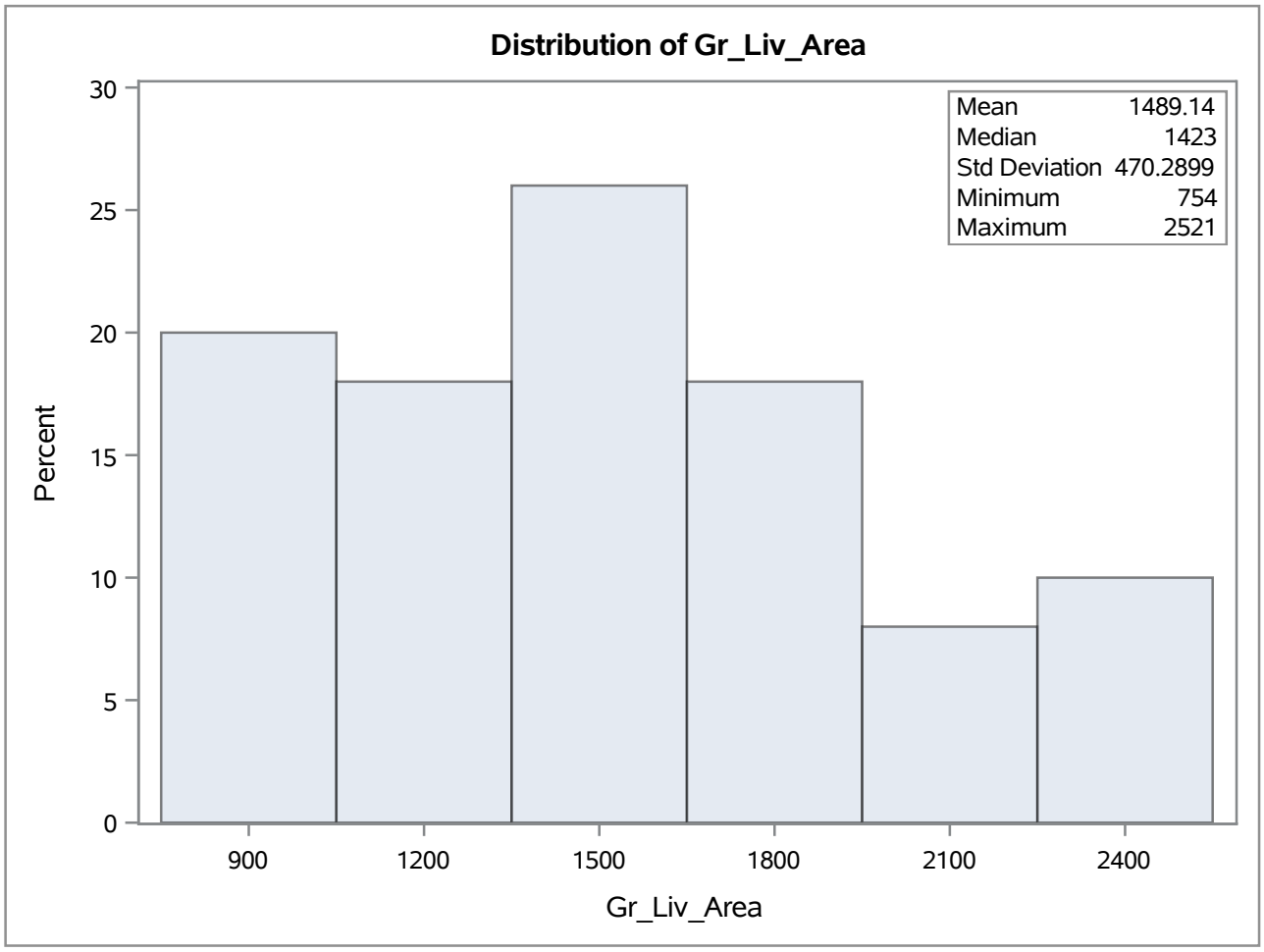
Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	22.39004	**Pr > |t|**	<.0001
Sign	**M**	25	**Pr >= |M|**	<.0001
Signed Rank	**S**	637.5	**Pr >= |S|**	<.0001

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	2521.0
99%	2521.0
95%	2461.0
90%	2211.5
75% Q3	1721.0
50% Median	1423.0
25% Q1	1131.0
10%	900.0
5%	844.0
1%	754.0
0% Min	754.0

The UNIVARIATE Procedure
Variable: Gr_Liv_Area

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
754	49	2262	32
833	11	2331	25
844	47	2461	4
874	10	2480	21
875	16	2521	50

The UNIVARIATE Procedure



The MEANS Procedure

Analysis Variable : Gr_Liv_Area
Mean
1489.14

Answer
The means of the two samples differ with the second one having a mean close to the population mean, while the first sample has a mean way higher of the population.

The SURVEYSELECT Procedure

Selection Method	Simple Random Sampling

Input Data Set	AMES
Random Number Seed	40334855
Sample Size	100
Selection Probability	0.03413
Sampling Weight	29.3
Output Data Set	AMESSAMPLE3

The MEANS Procedure

Analysis Variable : Gr_Liv_Area
Mean
1459.11

The SURVEYSELECT Procedure

Selection Method	Simple Random Sampling

Input Data Set	AMES
Random Number Seed	40362523
Sample Size	1000
Selection Probability	0.341297
Sampling Weight	2.93
Output Data Set	AMESSAMPLE4

The MEANS Procedure

Analysis Variable : Gr_Liv_Area
Mean
1505.72

Answer

After observing the mean of every sample we conclude that the higher the sample size is, the closer it gets to the population mean. So the sample of size 1000 provides a more accurate estimation.But, since all the samples are random each time, it is not always guaranteed that the bigger sample size will be more accurate.

The SURVEYSELECT Procedure

Selection Method	Simple Random Sampling

Input Data Set	AMES
Random Number Seed	40400812
Sample Size	50
Selection Probability	0.017065
Sampling Weight	58.6
Number of Replicates	5000
Total Sample Size	250000
Output Data Set	AMESSAMPLER

The UNIVARIATE Procedure
Variable: sampmean

Moments			
N	5000	**Sum Weights**	5000
Mean	1498.05339	**Sum Observations**	7490266.94
Std Deviation	70.5086375	**Variance**	4971.46796
Skewness	0.15671105	**Kurtosis**	0.31320389
Uncorrected SS	1.12457E10	**Corrected SS**	24852368.3
Coeff Variation	4.70668389	**Std Error Mean**	0.99714271

Basic Statistical Measures			
Location		**Variability**	
Mean	1498.053	**Std Deviation**	70.50864
Median	1497.560	**Variance**	4971
Mode	1475.500	**Range**	658.72000
		Interquartile Range	94.33000

Note: The mode displayed is the smallest of 2 modes with a count of 5.

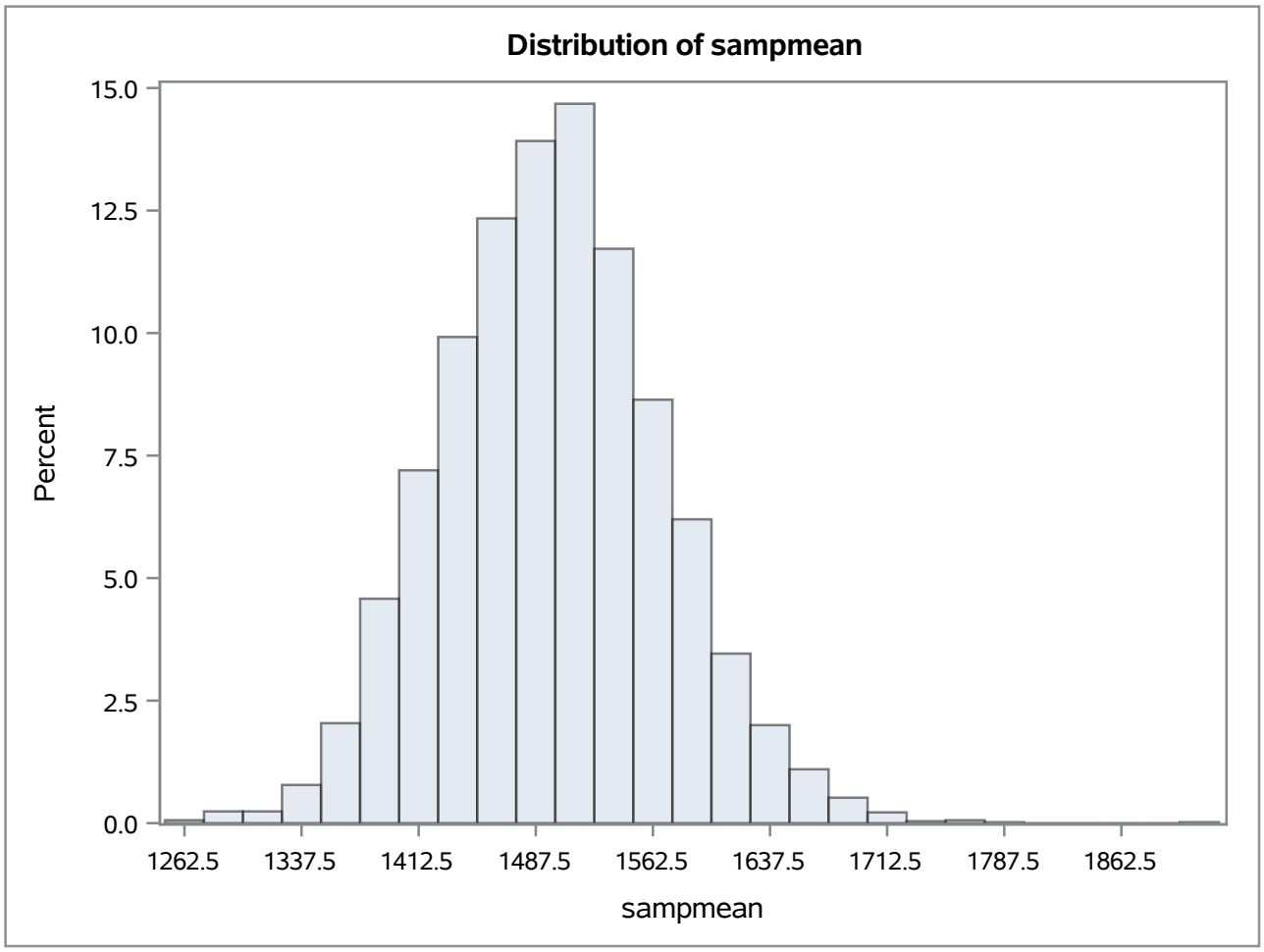
Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	1502.346	**Pr > |t|**	<.0001
Sign	**M**	2500	**Pr >= |M|**	<.0001
Signed Rank	**S**	6251250	**Pr >= |S|**	<.0001

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1910.82
99%	1671.82
95%	1615.81
90%	1588.60
75% Q3	1544.18
50% Median	1497.56
25% Q1	1449.85
10%	1408.69
5%	1385.38
1%	1339.84
0% Min	1252.10

The UNIVARIATE Procedure
Variable: sampmean

Extreme Observations			
Lowest		**Highest**	
Value	**Obs**	**Value**	**Obs**
1252.10	2425	1758.64	1903
1260.56	4669	1765.68	1393
1266.52	4744	1772.52	1156
1275.48	1116	1791.64	315
1276.66	3462	1910.82	1530

The UNIVARIATE Procedure



Answer

The work.reprum contains 5000 means with each mean originating from 50 samples, so the observations are 5000. We can observe from the histogram that it is almost symmetrical, so it is close to a normal distribution with the center being at 1500. The distribution would not really change at the 50.000 sample means, it will only get closer and closer to a normal distribution.

The SURVEYSELECT Procedure

Selection Method	Simple Random Sampling

Input Data Set	AMES
Random Number Seed	41760491
Sample Size	50
Selection Probability	0.017065
Sampling Weight	58.6
Number of Replicates	100
Total Sample Size	5000
Output Data Set	AMESSAMPLER2

Obs	Replicate	_TYPE_	_FREQ_	sampmean
1	1	0	50	1440.5
2	2	0	50	1579.1
3	3	0	50	1385.46
4	4	0	50	1455.84
5	5	0	50	1541.98
6	6	0	50	1557.64
7	7	0	50	1425.24
8	8	0	50	1578.62
9	9	0	50	1451.28
10	10	0	50	1426.66
11	11	0	50	1575.02
12	12	0	50	1618.48
13	13	0	50	1486.68
14	14	0	50	1505.06
15	15	0	50	1569.82
16	16	0	50	1610.2
17	17	0	50	1433.5
18	18	0	50	1433.2
19	19	0	50	1487.98
20	20	0	50	1569.34
21	21	0	50	1585.12
22	22	0	50	1567.56
23	23	0	50	1488.54
24	24	0	50	1471.52
25	25	0	50	1495.66
26	26	0	50	1356.68
27	27	0	50	1415.08
28	28	0	50	1616.9
29	29	0	50	1439.08
30	30	0	50	1493.06
31	31	0	50	1553
32	32	0	50	1409.72
33	33	0	50	1513.32
34	34	0	50	1399.64
35	35	0	50	1572.76
36	36	0	50	1450.06
37	37	0	50	1555.86
38	38	0	50	1492.42

Obs	Replicate	_TYPE_	_FREQ_	sampmean
39	39	0	50	1566.9
40	40	0	50	1566.94
41	41	0	50	1463.22
42	42	0	50	1421.74
43	43	0	50	1683.02
44	44	0	50	1630.42
45	45	0	50	1668.04
46	46	0	50	1381.1
47	47	0	50	1511.1
48	48	0	50	1432.32
49	49	0	50	1500.7
50	50	0	50	1397.62
51	51	0	50	1468.34
52	52	0	50	1307.9
53	53	0	50	1481.88
54	54	0	50	1395.5
55	55	0	50	1521.64
56	56	0	50	1418.88
57	57	0	50	1468.74
58	58	0	50	1644.4
59	59	0	50	1658.68
60	60	0	50	1515.46
61	61	0	50	1335.66
62	62	0	50	1549.52
63	63	0	50	1535.68
64	64	0	50	1488.12
65	65	0	50	1601.16
66	66	0	50	1519.46
67	67	0	50	1395.46
68	68	0	50	1598.62
69	69	0	50	1466.94
70	70	0	50	1648.36
71	71	0	50	1487.34
72	72	0	50	1521.3
73	73	0	50	1502.98
74	74	0	50	1530.62
75	75	0	50	1665.82
76	76	0	50	1621.16

Obs	Replicate	_TYPE_	_FREQ_	sampmean
77	77	0	50	1628.7
78	78	0	50	1487.24
79	79	0	50	1523.34
80	80	0	50	1596
81	81	0	50	1476.48
82	82	0	50	1390.28
83	83	0	50	1489.78
84	84	0	50	1416.44
85	85	0	50	1520.08
86	86	0	50	1436.34
87	87	0	50	1473.78
88	88	0	50	1542.72
89	89	0	50	1570.9
90	90	0	50	1476.2
91	91	0	50	1492.12
92	92	0	50	1593.74
93	93	0	50	1461.26
94	94	0	50	1465
95	95	0	50	1559.34
96	96	0	50	1490.5
97	97	0	50	1491.62
98	98	0	50	1588.82
99	99	0	50	1543.54
100	100	0	50	1448.86

Answer
The work.reprunsmall contains 100 observations. Each observation representing a sample mean from a sample of size 50.

The UNIVARIATE Procedure
Variable: sampmean

Moments			
N	5000	**Sum Weights**	5000
Mean	1498.05339	**Sum Observations**	7490266.94
Std Deviation	70.5086375	**Variance**	4971.46796
Skewness	0.15671105	**Kurtosis**	0.31320389
Uncorrected SS	1.12457E10	**Corrected SS**	24852368.3
Coeff Variation	4.70668389	**Std Error Mean**	0.99714271

Basic Statistical Measures			
Location		**Variability**	
Mean	1498.053	**Std Deviation**	70.50864
Median	1497.560	**Variance**	4971
Mode	1475.500	**Range**	658.72000
		Interquartile Range	94.33000

Note: The mode displayed is the smallest of 2 modes with a count of 5.

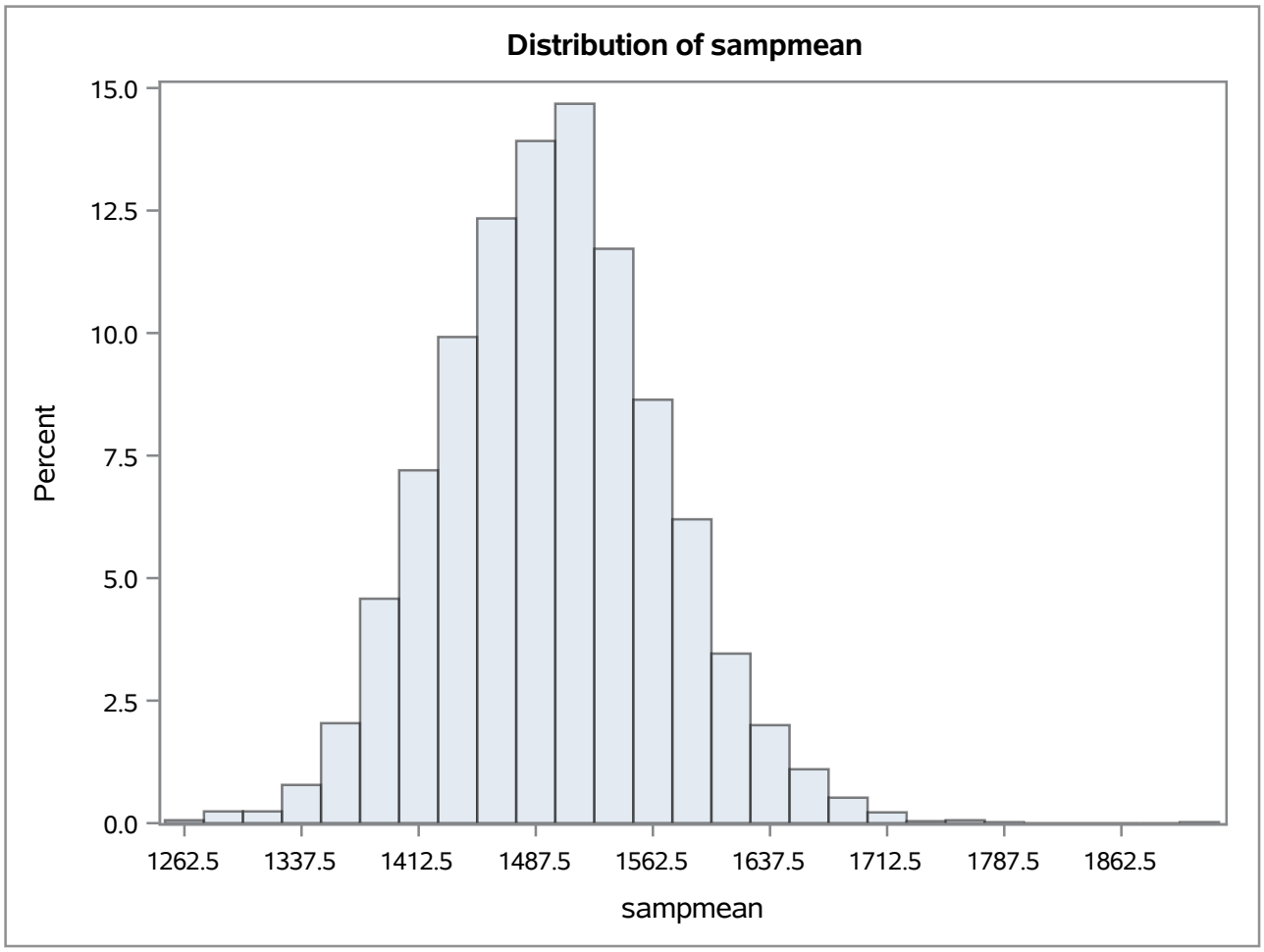
Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	1502.346	**Pr > |t|**	<.0001
Sign	**M**	2500	**Pr >= |M|**	<.0001
Signed Rank	**S**	6251250	**Pr >= |S|**	<.0001

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1910.82
99%	1671.82
95%	1615.81
90%	1588.60
75% Q3	1544.18
50% Median	1497.56
25% Q1	1449.85
10%	1408.69
5%	1385.38
1%	1339.84
0% Min	1252.10

The UNIVARIATE Procedure
Variable: sampmean

Extreme Observations			
Lowest		**Highest**	
Value	**Obs**	**Value**	**Obs**
1252.10	2425	1758.64	1903
1260.56	4669	1765.68	1393
1266.52	4744	1772.52	1156
1275.48	1116	1791.64	315
1276.66	3462	1910.82	1530

The UNIVARIATE Procedure



The SURVEYSELECT Procedure

Selection Method	Simple Random Sampling

Input Data Set	AMES
Random Number Seed	42137628
Sample Size	10
Selection Probability	0.003413
Sampling Weight	293
Number of Replicates	5000
Total Sample Size	50000
Output Data Set	AMESSAMPLER10

The SURVEYSELECT Procedure

Selection Method	Simple Random Sampling

Input Data Set	AMES
Random Number Seed	42693966
Sample Size	100
Selection Probability	0.03413
Sampling Weight	29.3
Number of Replicates	5000
Total Sample Size	500000
Output Data Set	AMESSAMPLER100

The UNIVARIATE Procedure
Variable: sampmean

Moments			
N	5000	**Sum Weights**	5000
Mean	1498.05339	**Sum Observations**	7490266.94
Std Deviation	70.5086375	**Variance**	4971.46796
Skewness	0.15671105	**Kurtosis**	0.31320389
Uncorrected SS	1.12457E10	**Corrected SS**	24852368.3
Coeff Variation	4.70668389	**Std Error Mean**	0.99714271

Basic Statistical Measures			
Location		**Variability**	
Mean	1498.053	**Std Deviation**	70.50864
Median	1497.560	**Variance**	4971
Mode	1475.500	**Range**	658.72000
		Interquartile Range	94.33000

Note: The mode displayed is the smallest of 2 modes with a count of 5.

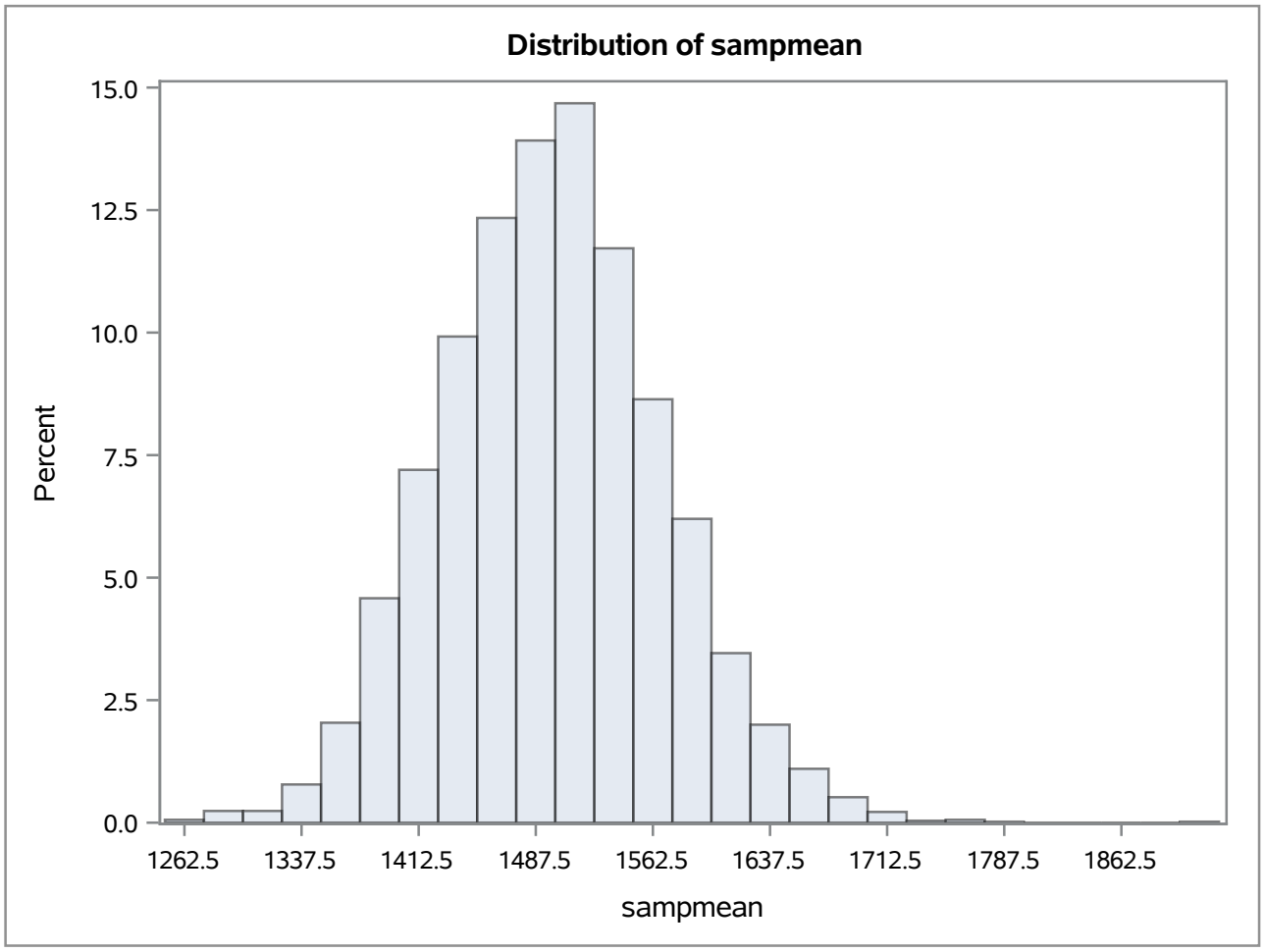
Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	1502.346	**Pr > |t|**	<.0001
Sign	**M**	2500	**Pr >= |M|**	<.0001
Signed Rank	**S**	6251250	**Pr >= |S|**	<.0001

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1910.82
99%	1671.82
95%	1615.81
90%	1588.60
75% Q3	1544.18
50% Median	1497.56
25% Q1	1449.85
10%	1408.69
5%	1385.38
1%	1339.84
0% Min	1252.10

The UNIVARIATE Procedure
Variable: sampmean

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
1252.10	2425	1758.64	1903
1260.56	4669	1765.68	1393
1266.52	4744	1772.52	1156
1275.48	1116	1791.64	315
1276.66	3462	1910.82	1530

The UNIVARIATE Procedure



The UNIVARIATE Procedure
Variable: sampmean

Moments			
N	5000	**Sum Weights**	5000
Mean	1500.37274	**Sum Observations**	7501863.7
Std Deviation	160.669965	**Variance**	25814.8376
Skewness	0.42214623	**Kurtosis**	0.54041945
Uncorrected SS	1.13846E10	**Corrected SS**	129048373
Coeff Variation	10.70867	**Std Error Mean**	2.27221643

Basic Statistical Measures			
Location		**Variability**	
Mean	1500.373	**Std Deviation**	160.66996
Median	1492.350	**Variance**	25815
Mode	1443.200	**Range**	1407
		Interquartile Range	209.20000

Note: The mode displayed is the smallest of 2 modes with a count of 7.

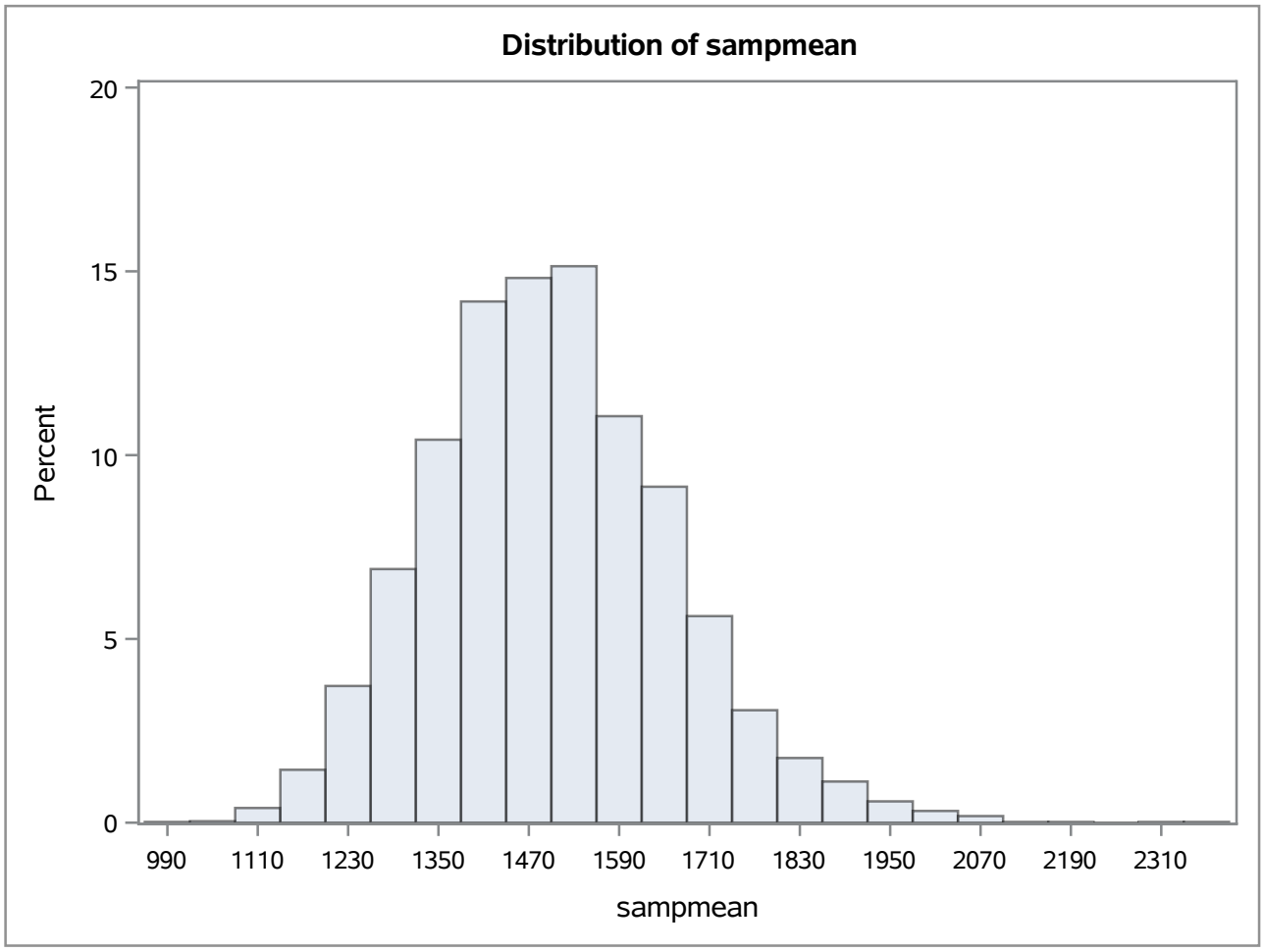
Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	660.3124	**Pr > |t|**	<.0001
Sign	**M**	2500	**Pr >= |M|**	<.0001
Signed Rank	**S**	6251250	**Pr >= |S|**	<.0001

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	2378.20
99%	1938.35
95%	1775.10
90%	1704.55
75% Q3	1600.40
50% Median	1492.35
25% Q1	1391.20
10%	1302.95
5%	1253.50
1%	1166.60
0% Min	971.20

The UNIVARIATE Procedure
Variable: sampmean

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
971.2	3318	2089.2	4597
1058.2	4889	2112.9	2676
1058.2	540	2181.9	2752
1083.9	24	2305.7	641
1084.7	27	2378.2	3293

The UNIVARIATE Procedure



The UNIVARIATE Procedure
Variable: sampmean

Moments			
N	5000	**Sum Weights**	5000
Mean	1499.52388	**Sum Observations**	7497619.4
Std Deviation	49.4445447	**Variance**	2444.763
Skewness	0.14193882	**Kurtosis**	0.08889784
Uncorrected SS	1.12551E10	**Corrected SS**	12221370.2
Coeff Variation	3.2973496	**Std Error Mean**	0.69925146

Basic Statistical Measures			
Location		**Variability**	
Mean	1499.524	**Std Deviation**	49.44454
Median	1498.525	**Variance**	2445
Mode	1514.850	**Range**	365.52000
		Interquartile Range	66.07500

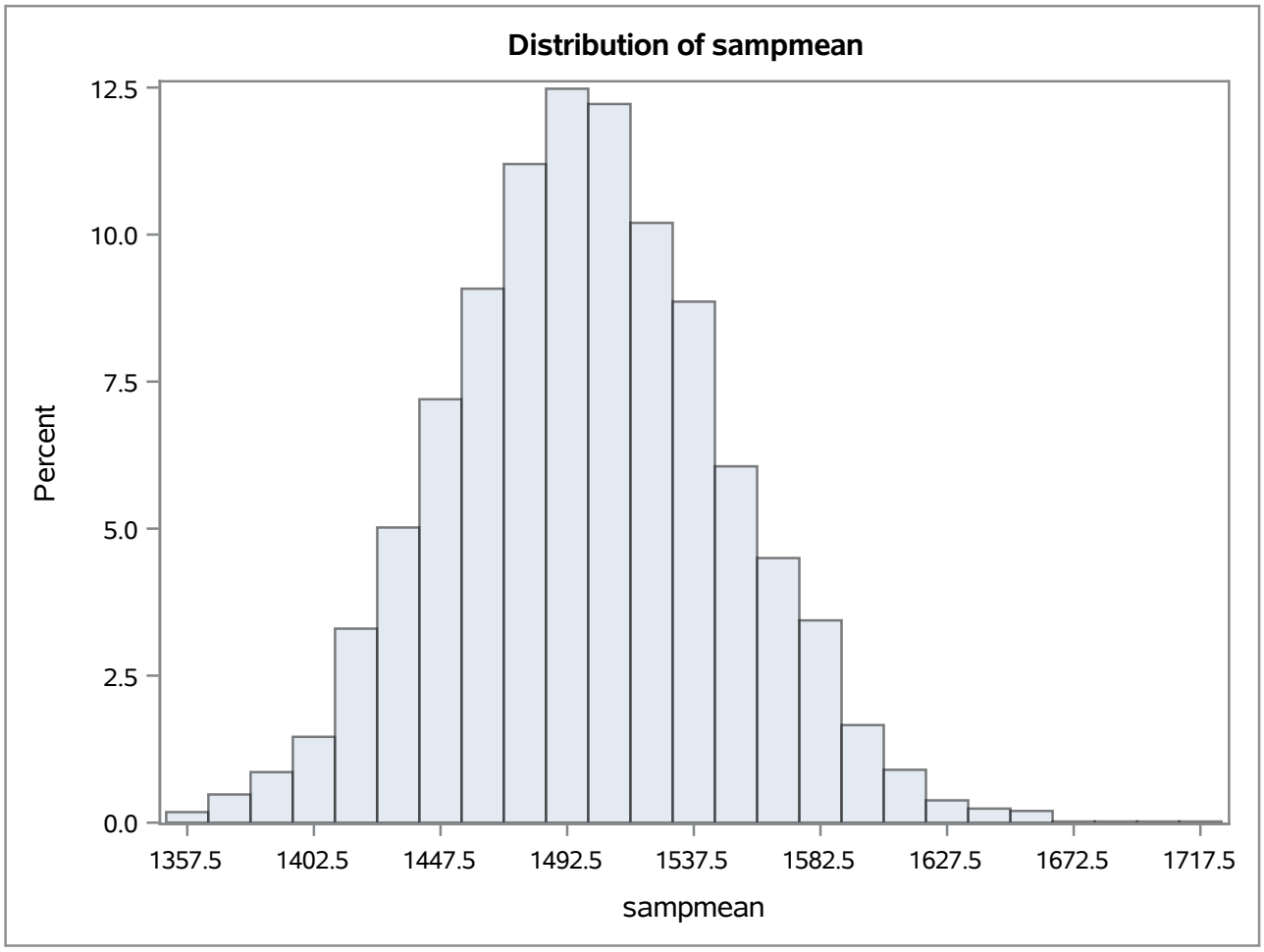
Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	2144.47	**Pr > |t|**	<.0001
Sign	**M**	2500	**Pr >= |M|**	<.0001
Signed Rank	**S**	6251250	**Pr >= |S|**	<.0001

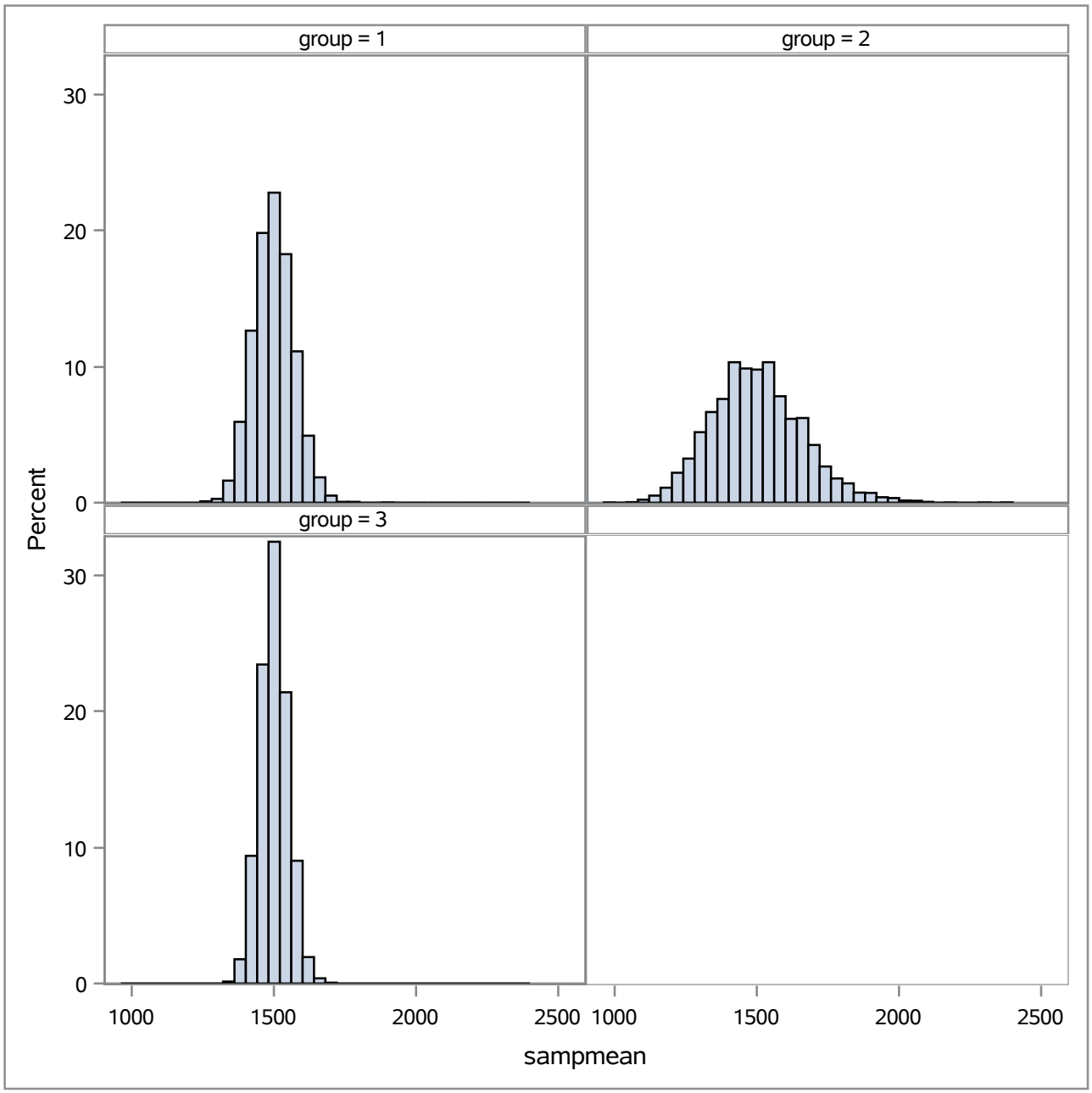
Quantiles (Definition 5)	
Level	**Quantile**
100% Max	1719.26
99%	1618.49
95%	1583.03
90%	1564.08
75% Q3	1532.14
50% Median	1498.53
25% Q1	1466.07
10%	1436.90
5%	1420.53
1%	1387.40
0% Min	1353.74

The UNIVARIATE Procedure
Variable: sampmean

Extreme Observations			
Lowest		**Highest**	
Value	**Obs**	**Value**	**Obs**
1353.74	4696	1661.48	4490
1354.90	4795	1670.48	2970
1355.74	4711	1686.06	2656
1357.28	144	1699.74	4148
1358.84	2869	1719.26	520

The UNIVARIATE Procedure




group = 1
group = 2
group = 3
Percent
30
20
10
0
1000
1500
2000
2500
sampmean

Answer
When the sample size is larger, the center still stays relatively the same in comparison to smaller sample sizes, but the center of the sampling disbribution would get closer and closer to the center of the population. However, the spread gets smaller as there is less variability.

1. Take a random sample of size 50 for SalePrice. Using this sample, what is your best point estimate of the population mean?.

The SURVEYSELECT Procedure

Selection Method	Simple Random Sampling

Input Data Set	AMES
Random Number Seed	72248513
Sample Size	50
Selection Probability	0.017065
Sampling Weight	58.6
Output Data Set	AMESSAMPLEPRICE

The MEANS Procedure

Analysis Variable : SalePrice
Mean
196307.44

2. Because you have access to the population, simulate the sampling distribution for $\bar{x}$ price by taking 5000 samples from the population of size 50 and computing 5000 sample means. Update the code to account for changing the variable in question. Name the storage data set work.repprice. Plot the data and then describe the shape of this sampling distribution. Based on this sampling distribution, what would you guess the mean home price of this population to be? Finally, calculate and report the population mean.

The SURVEYSELECT Procedure

Selection Method	Simple Random Sampling

Input Data Set	AMES
Random Number Seed	72400817
Sample Size	50
Selection Probability	0.017065
Sampling Weight	58.6
Number of Replicates	5000
Total Sample Size	250000
Output Data Set	AMESSAMPLEPRICE

The UNIVARIATE Procedure
Variable: sampmean

Moments			
N	5000	**Sum Weights**	5000
Mean	180833.68	**Sum Observations**	904168399
Std Deviation	11299.885	**Variance**	127687402
Skewness	0.25423577	**Kurtosis**	0.07925719
Uncorrected SS	1.64142E14	**Corrected SS**	6.38309E11
Coeff Variation	6.24877238	**Std Error Mean**	159.804507

Basic Statistical Measures			
Location		**Variability**	
Mean	180833.7	**Std Deviation**	11300
Median	180442.8	**Variance**	127687402
Mode	162669.6	**Range**	81434
		Interquartile Range	15062

Note: The mode displayed is the smallest of 10 modes with a count of 2.

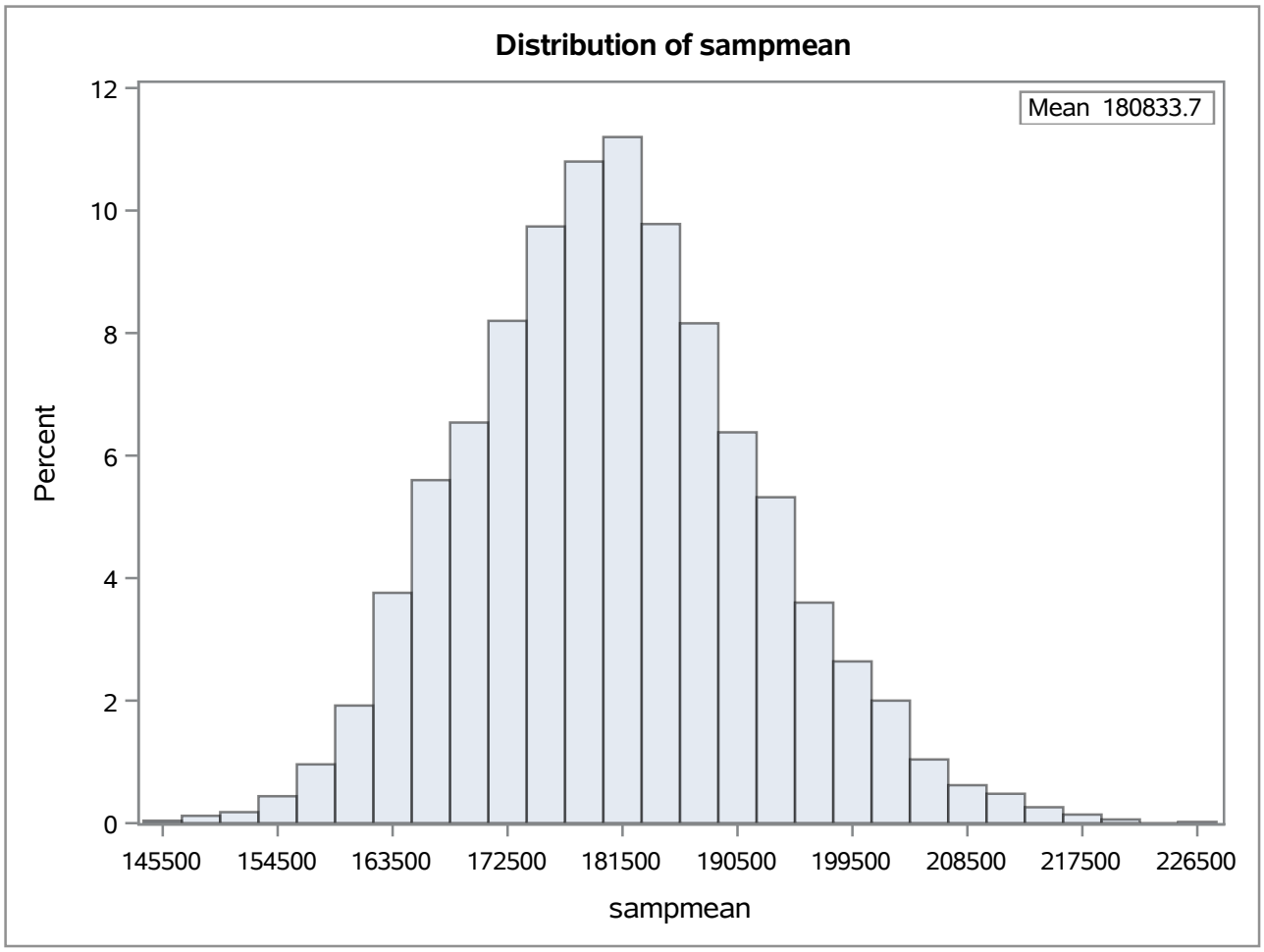
Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	1131.593	**Pr > |t|**	<.0001
Sign	**M**	2500	**Pr >= |M|**	<.0001
Signed Rank	**S**	6251250	**Pr >= |S|**	<.0001

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	227055
99%	209764
95%	200498
90%	195544
75% Q3	188047
50% Median	180443
25% Q1	172985
10%	166513
5%	162996
1%	156941
0% Min	145621

The UNIVARIATE Procedure
Variable: sampmean

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
145621	2722	217907	2530
146901	2678	219131	677
148177	192	220563	3787
148429	1097	221819	1321
148806	4955	227055	1604

The UNIVARIATE Procedure



The MEANS Procedure

Analysis Variable : SalePrice
Mean
180796.06

Description
The shape of the sampling mean distribution resembles a normal distrubtion since we have a high sample size, the sample mean home price is at 180.700 which can differ but with minimal deviation since it originates from 5000 random samples of size 50. The spread is small. Before calculating the population mean home price I am estiamting it to be around 180.000, which is very close to the sampling mean.

3. Change your sample size from 50 to 150, and then compute the sampling distribution using the same method as above. Store these means in a new data set named work.repprice150. Describe the shape of this sampling distribution, and compare it to the sampling distribution for a sample size of 50. Based on this sampling distribution, what would you guess to be the mean sale price of homes in Ames?

The SURVEYSELECT Procedure

Selection Method	Simple Random Sampling

Input Data Set	AMES
Random Number Seed	75250662
Sample Size	150
Selection Probability	0.051195
Sampling Weight	19.53333
Number of Replicates	5000
Total Sample Size	750000
Output Data Set	AMESSAMPLEPRICE2

The UNIVARIATE Procedure
Variable: sampmean

Moments			
N	5000	**Sum Weights**	5000
Mean	180808.265	**Sum Observations**	904041327
Std Deviation	6411.2905	**Variance**	41104645.8
Skewness	0.16395361	**Kurtosis**	-0.02634
Uncorrected SS	1.63664E14	**Corrected SS**	2.05482E11
Coeff Variation	3.54590565	**Std Error Mean**	90.6693397

Basic Statistical Measures			
Location		**Variability**	
Mean	180808.3	**Std Deviation**	6411
Median	180608.0	**Variance**	41104646
Mode	173089.3	**Range**	44957
		Interquartile Range	8674

Note: The mode displayed is the smallest of 2 modes with a count of 2.

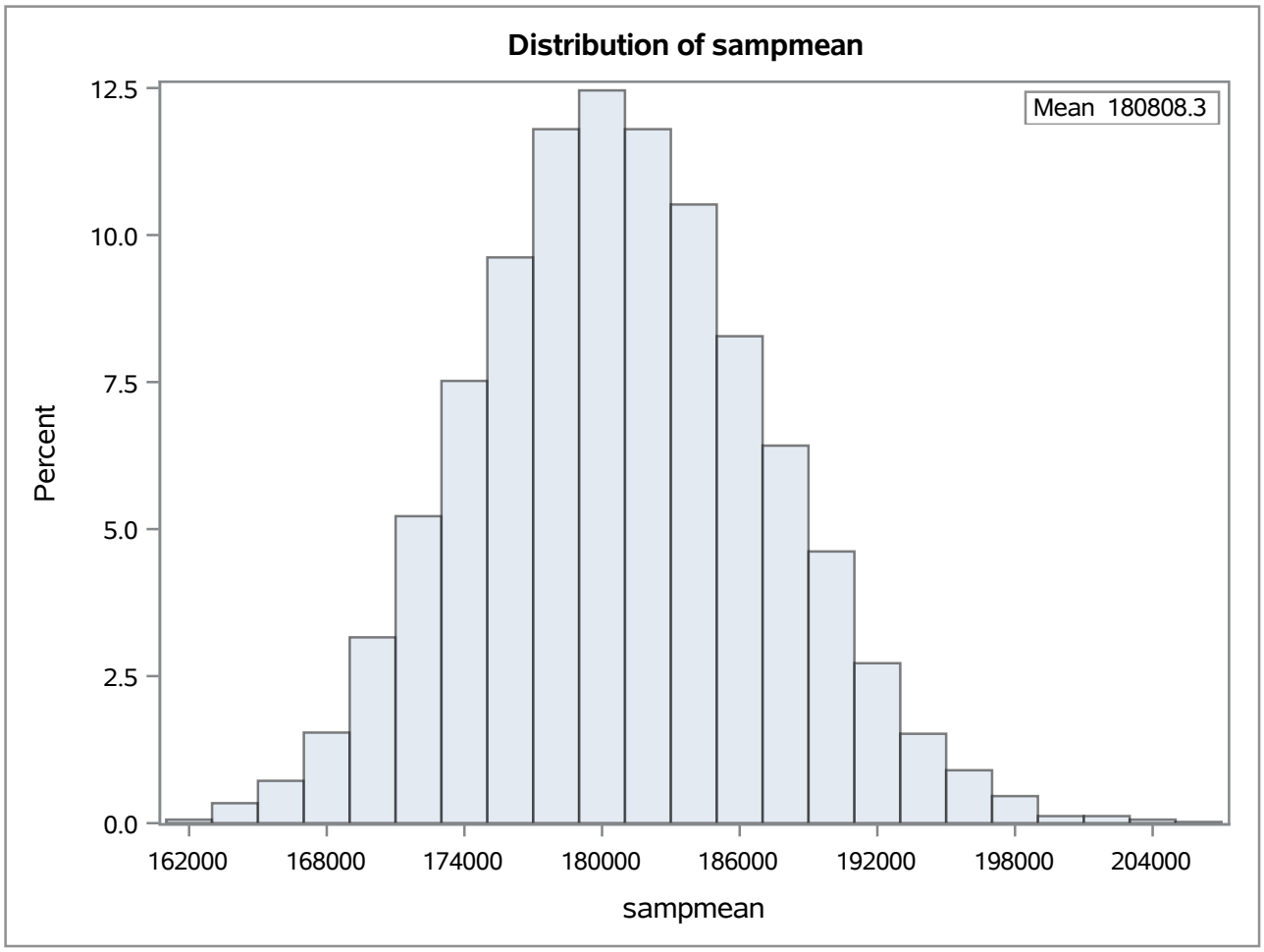
Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	1994.15	**Pr > |t|**	<.0001
Sign	**M**	2500	**Pr >= |M|**	<.0001
Signed Rank	**S**	6251250	**Pr >= |S|**	<.0001

Quantiles (Definition 5)	
Level	**Quantile**
100% Max	206283
99%	196343
95%	191495
90%	189171
75% Q3	185026
50% Median	180608
25% Q1	176352
10%	172588
5%	170551
1%	166648
0% Min	161327

The UNIVARIATE Procedure
Variable: sampmean

Extreme Observations			
Lowest		**Highest**	
Value	**Obs**	**Value**	**Obs**
161327	2689	202852	1582
162164	4213	203259	1691
162699	4094	203327	837
163161	3612	204374	1729
163336	4198	206283	2534

The UNIVARIATE Procedure



Description
The shape of the 150 sampling mean distribution still resembles a normal distrubtion and it is closer to that distribution compared to the sample distribution of 50 samples. Also, the sample mean home price is closer to the population mean which also can differ but with a std deviation lower than the sample distribution of 50 we have higher accuracy. The spread is smaller. My guess would still be around 180.000.

4. Of the sampling distributions from 2 and 3, which has a smaller spread? If we're concerned with making estimates that are more often close to the true value, would we prefer a distribution with a large or small spread?.

Description
After observing the two sampling distributions, the distribution from 3 has smaller spread. To estimate closer to the true value we want a distribution with a large spread.