Comparing two sample averages

Gaston Sanchez

Preliminary concepts

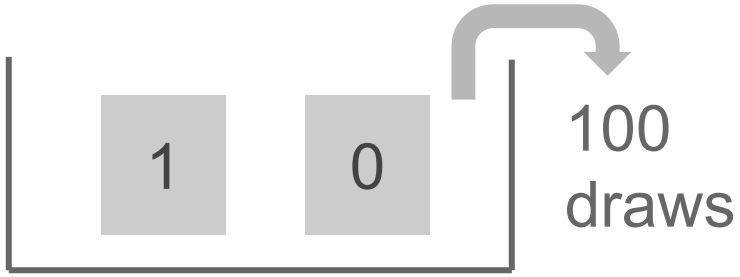
(FPP chapter 27)

Motivation

Toss a coin 100 times, count # heads

Then toss coin 100 more times, count # heads

Consider (# draws in 1st set) + (# draws in 2nd set), the expected value should be around ___100___ give or take ___?___

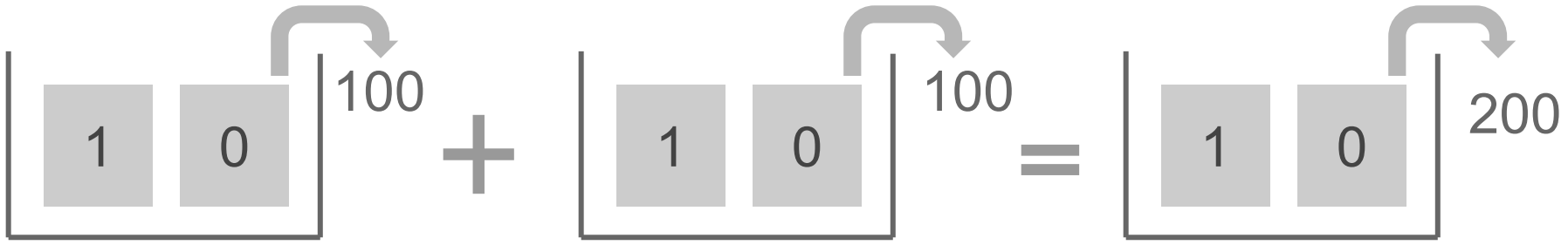


Avg of box = (1 + 0) / 2 = 0.5

EV of sum = 100 x 0.5 = 50

SD of box = $(1-0)\sqrt{(0.5)(0.5)} = 0.5$

SE sum = $\sqrt{100}\,(0.5) = 5$



Avg of box = (1 + 0) / 2 = 0.5

EV of sum = 200 x 0.5 = 100

SD of box = $(1-0)\sqrt{(0.5)(0.5)} = 0.5$

SE sum = $\sqrt{200}\,(0.5) = 5\sqrt{2} = 7.07$

Motivation

Toss a coin 100 times, count # heads

Then toss coin 100 more times, count # heads

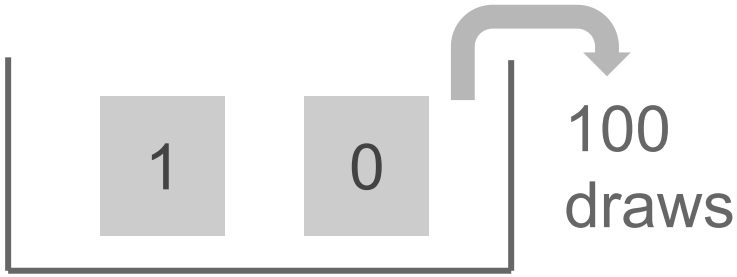
Consider (# draws in 1st set) + (# draws in 2nd set), the expected value should be around ***100*** give or take ***7.07***

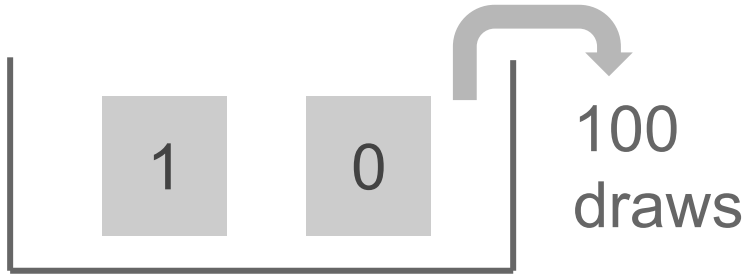
Another example

Subtracting sum of draws

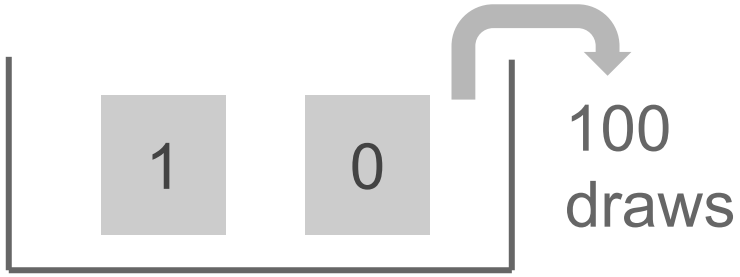
Toss a coin 100 times, count # heads

Then toss coin 100 more times, count # heads





Consider (# draws in 1st set) - (# draws in 2nd set), the expected value should be around ___*0*___ give or take ___*?*___

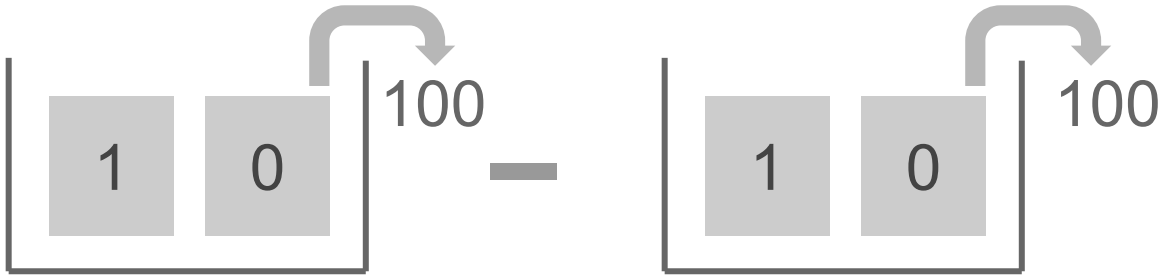


Avg of box = (1 + 0) / 2 = 0.5

EV of sum = 100 x 0.5 = 50

SD of box = $(1-0)\sqrt{(0.5)(0.5)} = 0.5$

SE sum = $\sqrt{100}\,(0.5) = 5$



Avg of box = (1 + 0) / 2 = 0.5

EV of subtraction = 50 - 50 = 0

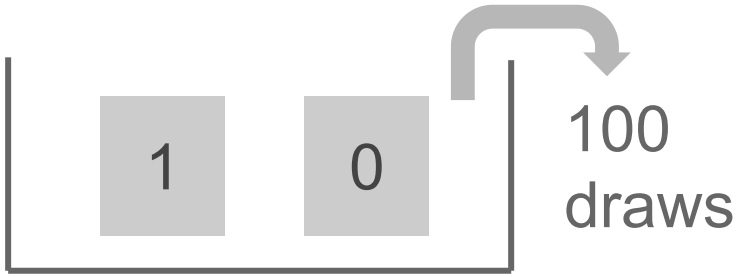
SD of box = $(1-0)\sqrt{(0.5)(0.5)} = 0.5$

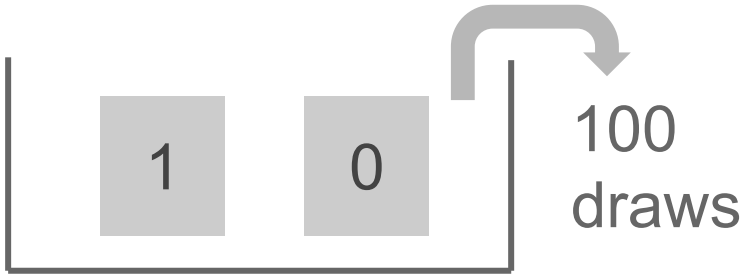
SE sum = $\sqrt{200}\,(0.5) = 5\sqrt{2} = 7.07$

Motivation

Toss a coin 100 times, count # heads

Then toss coin 100 more times, count # heads





Consider (# draws in 1st set) - (# draws in 2nd set), the expected value should be around ___*0*___ give or take ___*7.07*___

Two sample z-test

Two sample z-test used for

1) Comparing values (sums, %'s, avgs) from two independent SRS (or samples w/replacement)

2) Comparing values from two groups (e.g. treatment and control) both selected randomly from one group.

Two sample z-test

1) Two independent SRS (or with replacement)

2) One population divided randomly into treatment and control groups (not necessarily of the same size)

Warriors Example


GOLDEN STATE
WARRIORS

Regular Season 2014-2015

Regular Season Glossary · SHARE · Embed · CSV · Export · PRE · LINK · ?

G	Date					Opponent			Tm	Opp	W	L	Streak
1	Wed, Oct 29, 2014	10:00p EST		Box Score	@	Sacramento Kings	W		95	77	1	0	W 1
2	Sat, Nov 1, 2014	10:30p EST		Box Score		Los Angeles Lakers	W		127	104	2	0	W 2
3	Sun, Nov 2, 2014	9:00p EST		Box Score	@	Portland Trail Blazers	W		95	90	3	0	W 3
4	Wed, Nov 5, 2014	10:30p EST		Box Score		Los Angeles Clippers	W		121	104	4	0	W 4
5	Sat, Nov 8, 2014	8:00p EST		Box Score	@	Houston Rockets	W		98	87	5	0	W 5
6	Sun, Nov 9, 2014	8:00p EST		Box Score	@	Phoenix Suns	L		95	107	5	1	L 1
7	Tue, Nov 11, 2014	10:30p EST		Box Score		San Antonio Spurs	L		100	113	5	2	L 2
8	Thu, Nov 13, 2014	10:30p EST		Box Score		Brooklyn Nets	W		107	99	6	2	W 1
9	Sat, Nov 15, 2014	10:30p EST		Box Score		Charlotte Hornets	W		112	87	7	2	W 2
10	Sun, Nov 16, 2014	9:30p EST		Box Score	@	Los Angeles Lakers	W		136	115	8	2	W 3
11	Fri, Nov 21, 2014	10:30p EST		Box Score		Utah Jazz	W		101	88	9	2	W 4
12	Sun, Nov 23, 2014	7:00p EST		Box Score	@	Oklahoma City Thunder	W		91	86	10	2	W 5
13	Tue, Nov 25, 2014	7:30p EST		Box Score	@	Miami Heat	W		114	97	11	2	W 6
14	Wed, Nov 26, 2014	7:00p EST		Box Score	@	Orlando Magic	W		111	96	12	2	W 7
15	Fri, Nov 28, 2014	7:00p EST		Box Score	@	Charlotte Hornets	W		106	101	13	2	W 8

Regular Season 2015-2016

Regular Season Glossary · SHARE · Embed · CSV · Export · PRE · LINK · ?

G	Date				Opponent			Tm	Opp	W	L	Streak
1	Tue, Oct 27, 2015	10:30p EST	Box Score		New Orleans Pelicans	W		111	95	1	0	W 1
2	Fri, Oct 30, 2015	9:30p EST	Box Score	@	Houston Rockets	W		112	92	2	0	W 2
3	Sat, Oct 31, 2015	7:30p EST	Box Score	@	New Orleans Pelicans	W		134	120	3	0	W 3
4	Mon, Nov 2, 2015	10:30p EST	Box Score		Memphis Grizzlies	W		119	69	4	0	W 4
5	Wed, Nov 4, 2015	10:30p EST	Box Score		Los Angeles Clippers	W		112	108	5	0	W 5
6	Fri, Nov 6, 2015	10:30p EST	Box Score		Denver Nuggets	W		119	104	6	0	W 6
7	Sat, Nov 7, 2015	10:00p EST	Box Score	@	Sacramento Kings	W		103	94	7	0	W 7
8	Mon, Nov 9, 2015	10:30p EST	Box Score		Detroit Pistons	W		109	95	8	0	W 8
9	Wed, Nov 11, 2015	8:00p EST	Box Score	@	Memphis Grizzlies	W		100	84	9	0	W 9
10	Thu, Nov 12, 2015	8:00p EST	Box Score	@	Minnesota Timberwolves	W		129	116	10	0	W 10
11	Sat, Nov 14, 2015	10:30p EST	Box Score		Brooklyn Nets	W	OT	107	99	11	0	W 11
12	Tue, Nov 17, 2015	10:30p EST	Box Score		Toronto Raptors	W		115	110	12	0	W 12
13	Thu, Nov 19, 2015	10:30p EST	Box Score	@	Los Angeles Clippers	W		124	117	13	0	W 13
14	Fri, Nov 20, 2015	10:30p EST	Box Score		Chicago Bulls	W		106	94	14	0	W 14
15	Sun, Nov 22, 2015	8:00p EST	Box Score	@	Denver Nuggets	W		118	105	15	0	W 15

SRS of 50 games

Season 2015

Games = 50

Sample Avg = 109.5

Sample SD = 11.57

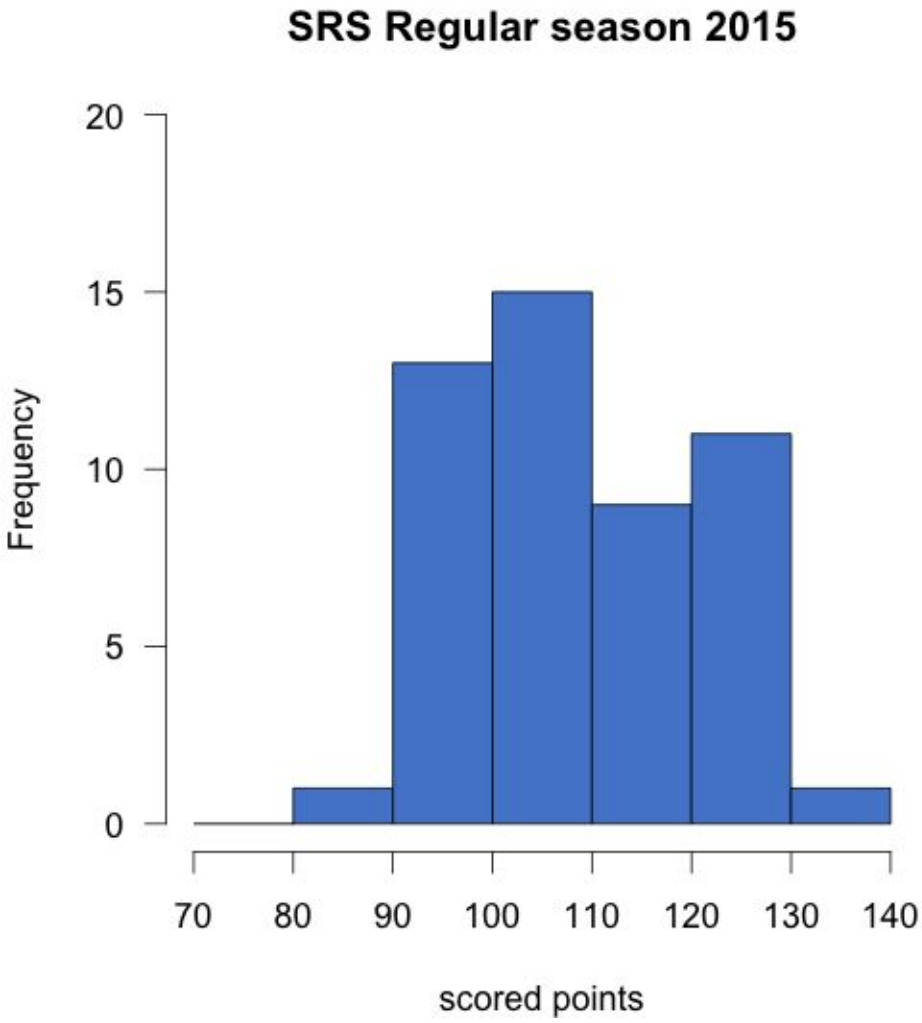
Season 2016

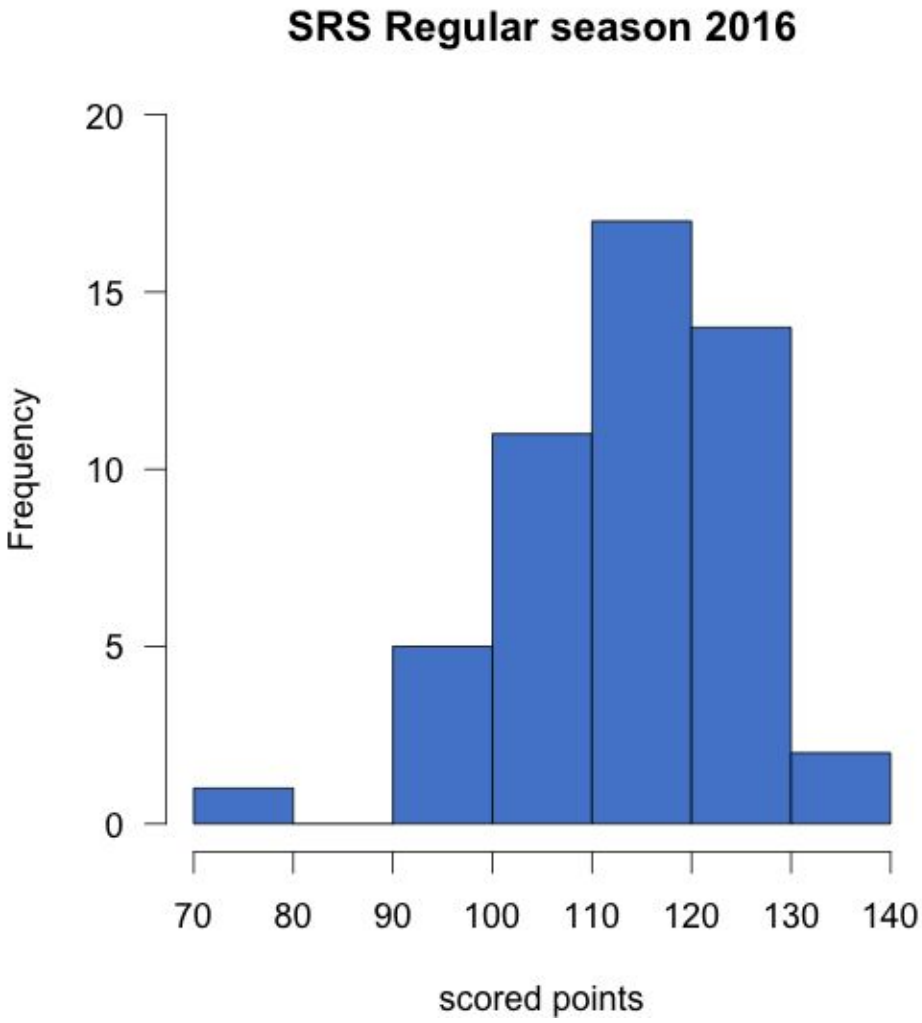
Games = 50

Sample Avg = 114.18

Sample SD = 11.97

Note: If the sample size n is large, sample SD is very likely to be close to SD(box)


SRS Regular season 2015
Frequency
0
5
10
15
20
70
80
90
100
110
120
130
140
scored points


SRS Regular season 2016
Frequency
0
5
10
15
20
70
80
90
100
110
120
130
140
scored points

Observed difference:
109.5 -vs- 114.18

Is this difference real?

Or can it be explained by chance?

Warriors scores

Null:

a) The difference 109.5 -vs- 114.18 can be explained by chance
b) The two avgs should be the same

Alternative:

a) Not just random chance
b) Avg season 2016 > Avg season 2015

Two sample z-test

Test statistic:

$$Z = \frac{\text{Obs. difference} \; - \; \text{Exp. difference}}{\text{SE difference}}$$

$$Z = \frac{(109.5 - 114.18) \; - \; (0)}{\text{SE difference}}$$

SE Season 2015

Sample SD “2015” = 11.75

SE avg “2015” = $11.75 / \sqrt{50} = 1.63$

SE Season 2016

Sample SD “2016” = 11.97

SE sum “2016” = $11.97 / \sqrt{50} = 1.69$

Z-test

$$\text{SE diff} = \sqrt{(1.63)^2 + (1.69)^2} = 2.34$$

$$Z = \frac{(109.5 - 114.18) - (0)}{2.34} = -2$$

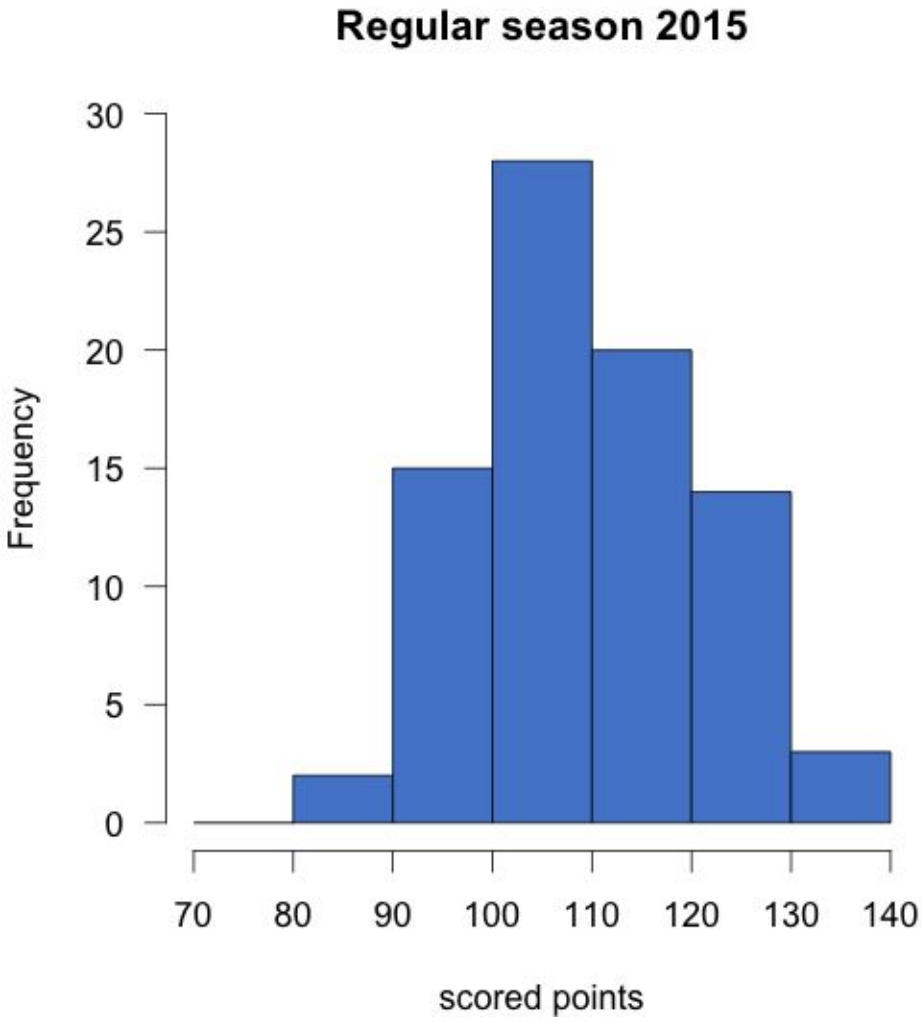
$$\text{P-value} = 0.022 < 0.05$$

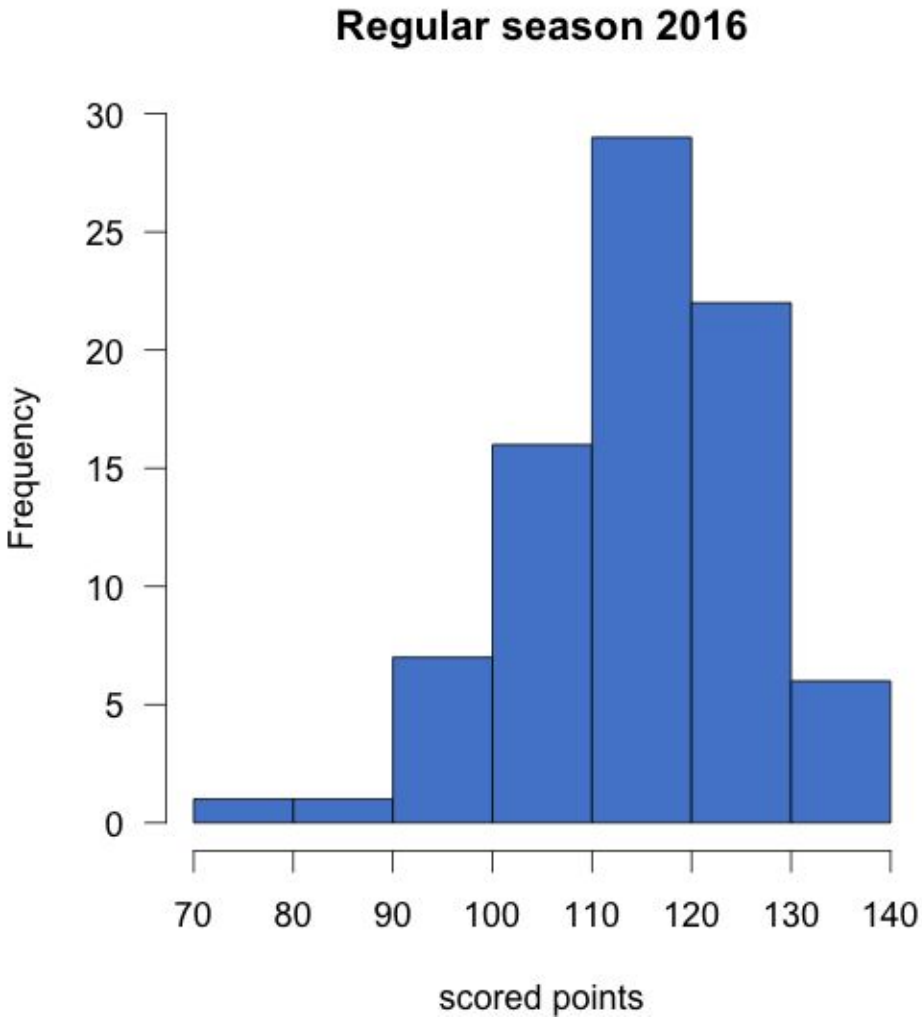
Warriors scores

Conclusion:

- At a 5% significance level, we reject the null hypothesis.
- The difference (109.5 -vs- 114.18) doesn't seem to be explained by chance.
- Avg score for season 2016 is greater than avg score for season 2015

Actually ....


Regular season 2015
Frequency
0
5
10
15
20
25
30
70
80
90
100
110
120
130
140
scored points


Regular season 2016
Frequency
0
5
10
15
20
25
30
70
80
90
100
110
120
130
140
scored points

Warriors' Scores

Season 2015

Games = 82

Avg = 110

SD = 10.87

Season 2016

Games = 82

Avg = 114.9

SD = 11.62

Comparing two percentages

You can also use the two sample z-test to compare two percentages


STRENGTH IN NUMBERS
30

Curry's three point percentage efficacy

Season 2015

Attempts = 646

Made = 286

3P% = **0.443**

Season 2016

Attempts = 886

Made = 402

3P% = **0.454**

Observed difference:

44.3% -vs- **45.4%**

Is this difference real?

Or can it be explained by chance?

Curry's 3P%

Null:

a) The difference **44.3%** -vs- **45.4%** can be explained by chance
b) 3P% 2015 = 3P% 2016

Alternative:

a) Not just random chance
b) 3P% 2016 > 3P% 2015

Two sample z-test

Test statistic:

$$Z = \frac{\text{Obs. difference} \;-\; \text{Exp. difference}}{\text{SE difference}}$$

$$Z = \frac{(0.443 - 0.454) \;-\; (0)}{\text{SE difference}}$$

Season 2015

$$\text{SD sample} = \sqrt{0.443 \times 0.557} = 0.4967$$

$$\text{SE \%} = 0.4967 / \sqrt{646} = 1.9$$

Season 2016

$$\text{SD sample} = \sqrt{0.454 \times 0.546} = 0.4978$$

$$\text{SE \%} = 0.4978 / \sqrt{886} = 1.6\%$$

Z-test

$$\text{SE diff} = \sqrt{(1.6)^2 + (1.9)^2} = 2.48$$

$$Z = \frac{(44.3 - 45.4) \; - \; (0)}{2.48} = -0.44$$

$$\text{P-value} = 0.32 \; > \; 0.05$$

Curry's 3P%

Conclusion:

- At a 5% significance level, we fail to reject the null hypothesis.
- The difference (44.3% -vs- 45.4%) seems to be explained by chance.
- 3P% season 2016 is not really different from 3P% season 2015