Randomization Tests - Beyond One/Two Sample Means & Proportions

Patti Frazer Lock, St. Lawrence University

Robin Lock, St. Lawrence University

Kari Lock Morgan, Penn State

USCOTS 2019

Randomization Test

Basic Procedure:

1. Calculate a test statistic for the original sample.
2. Simulate a new (randomization) sample under the null hypothesis.
3. Calculate the test statistic for the new sample.
4. Repeat 2 & 3 thousands of times to generate a randomization distribution.
5. Find a p-value as the proportion of simulated samples that give a test statistic as (or more) extreme as the original sample.

Tests in this Breakout

Chi-square goodness-of-fit

Chi-square test for association	Cat. vs. Cat.
ANOVA for means	Cat. vs. Quant.
ANOVA for regression	Quant. vs. Quant.

These all test for a relationship

H_0: No relationship

How do we use the data to simulate samples under this null hypothesis?

No Relationship via Scrambling

x_1	y_1
x_2	y_2
x_3	y_3
x_4	y_4
x_5	y_5
x_6	y_6
x_7	y_7
x_8	y_8
x_9	y_9

y_1
y_2
y_3
y_4
y_5
y_6
y_7
y_8
y_9

Two Quantitative

No Relationship via Scrambling

x_1	y_8
x_2	y_7
x_3	y_5
x_4	y_4
x_5	y_1
x_6	y_6
x_7	y_9
x_8	y_2
x_9	y_3

y_1
y_2
y_3
y_4
y_5
y_6
y_7
y_8
y_9

Two Quantitative

No Relationship via Scrambling

x_1	y_1
x_2	y_2
x_3	y_3
x_4	y_4
x_5	y_5
x_6	y_6
x_7	y_7
x_8	y_8
x_9	y_9

Two Quantitative

A	y_1
A	y_2
A	y_3
B	y_4
B	y_5
B	y_6
C	y_7
C	y_8
C	y_9

One Categorical
One Quantitative

y_1
y_2
y_3
y_4
y_5
y_6
y_7
y_8
y_9

No Relationship via Scrambling

x_1	y_8
x_2	y_7
x_3	y_5
x_4	y_4
x_5	y_1
x_6	y_6
x_7	y_9
x_8	y_2
x_9	y_3

Two Quantitative

A	y_8
A	y_7
A	y_5
B	y_4
B	y_1
B	y_6
C	y_9
C	y_2
C	y_3

One Categorical

One Quantitative

y_1
y_2
y_3
y_4
y_5
y_6
y_7
y_8
y_9

No Relationship via Scrambling

x_1	y_1
x_2	y_2
x_3	y_3
x_4	y_4
x_5	y_5
x_6	y_6
x_7	y_7
x_8	y_8
x_9	y_9

Two Quantitative

A	y_1
A	y_2
A	y_3
B	y_4
B	y_5
B	y_6
C	y_7
C	y_8
C	y_9

One Categorical
One Quantitative

A	*yes*
A	*no*
A	*no*
B	*yes*
B	*no*
B	*yes*
C	*yes*
C	*yes*
C	*no*

yes
no
no
yes
no
yes
yes
yes
no

Two Categorical

No Relationship via Scrambling

x_1	y_8
x_2	y_7
x_3	y_5
x_4	y_4
x_5	y_1
x_6	y_6
x_7	y_9
x_8	y_2
x_9	y_3

Two Quantitative

A	y_8
A	y_7
A	y_5
B	y_4
B	y_1
B	y_6
C	y_9
C	y_2
C	y_3

One Categorical
One Quantitative

A	yes
A	yes
A	no
B	yes
B	yes
B	yes
C	no
C	no
C	no

yes
no
no
yes
no
yes
yes
yes
no

Two Categorical

What Statistic?

We can scramble to simulate samples under a null of "no relationship". What statistic should we compute for each sample?

Chi-square for Association:

$$\chi^2 = \sum \frac{(observed - expected)^2}{expected}$$

ANOVA for Means:

$$F = \frac{MSG}{MSE} = \frac{\sum n_i(\bar{x}_i - \bar{x})^2/df_1}{\sum (x - \bar{x}_i)^2/df_2}$$

Let technology take care of calculations

ANOVA for Regression:

$$F = \frac{MSModel}{MSE} = \frac{\sum (\hat{y} - \bar{y})^2/df_1}{\sum (y - \hat{y})^2/df_2}$$

Example #1: Which Award?

If you could win an Olympic Gold Medal, Academy Award, or Nobel Prize, which would you choose?

Do think the distributions will differ between male and female students?

	Olympic	Academy	Nobel	
Male	109 (97.0)	11 (16.5)	73 (79.4)	193
Female	73 (85.0)	20 (14.5)	76 (69.6)	169
	182	31	149	n=362

$\chi^2 = 8.24$ Is that an unusually large value?

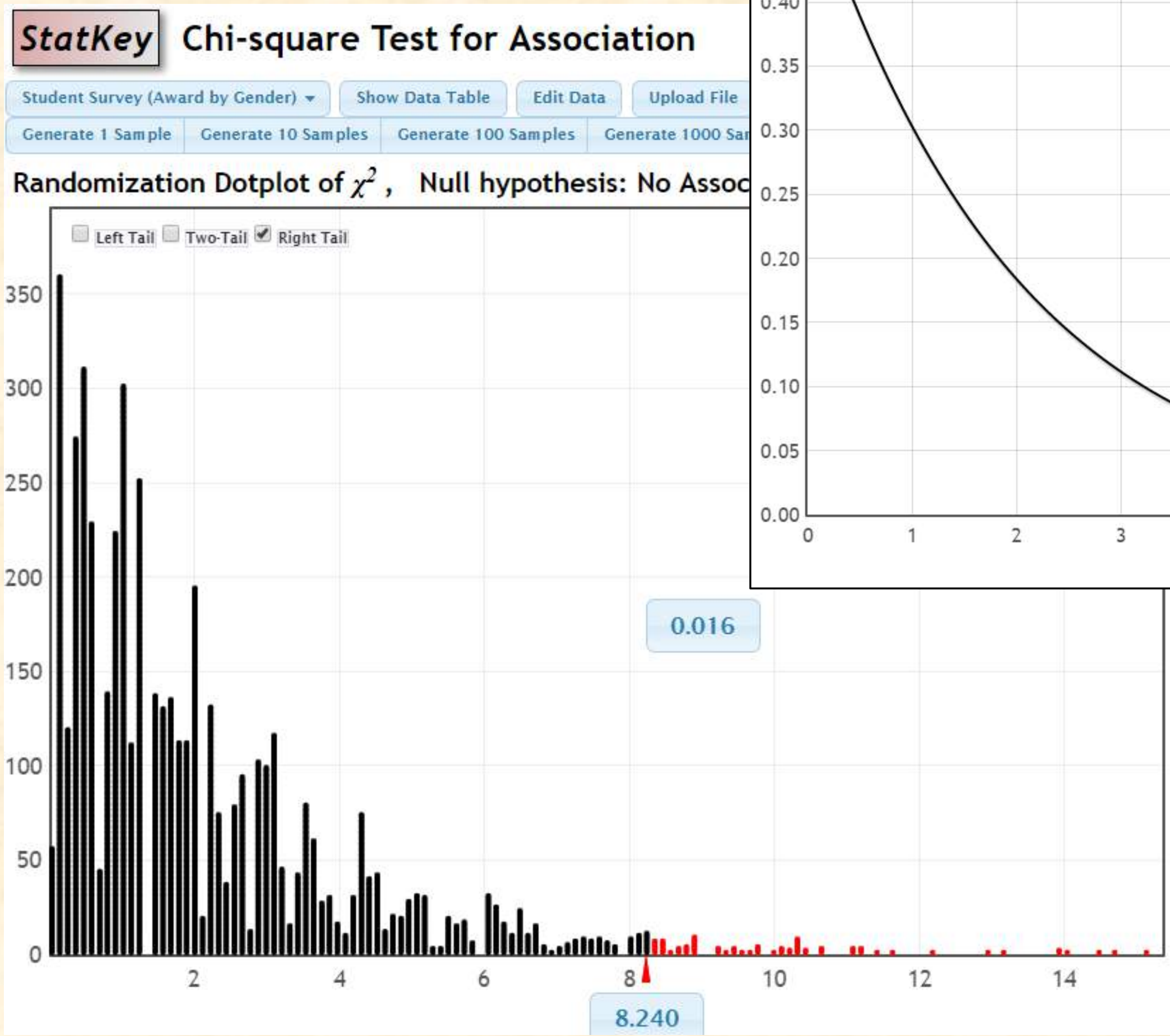
Randomization for Awards

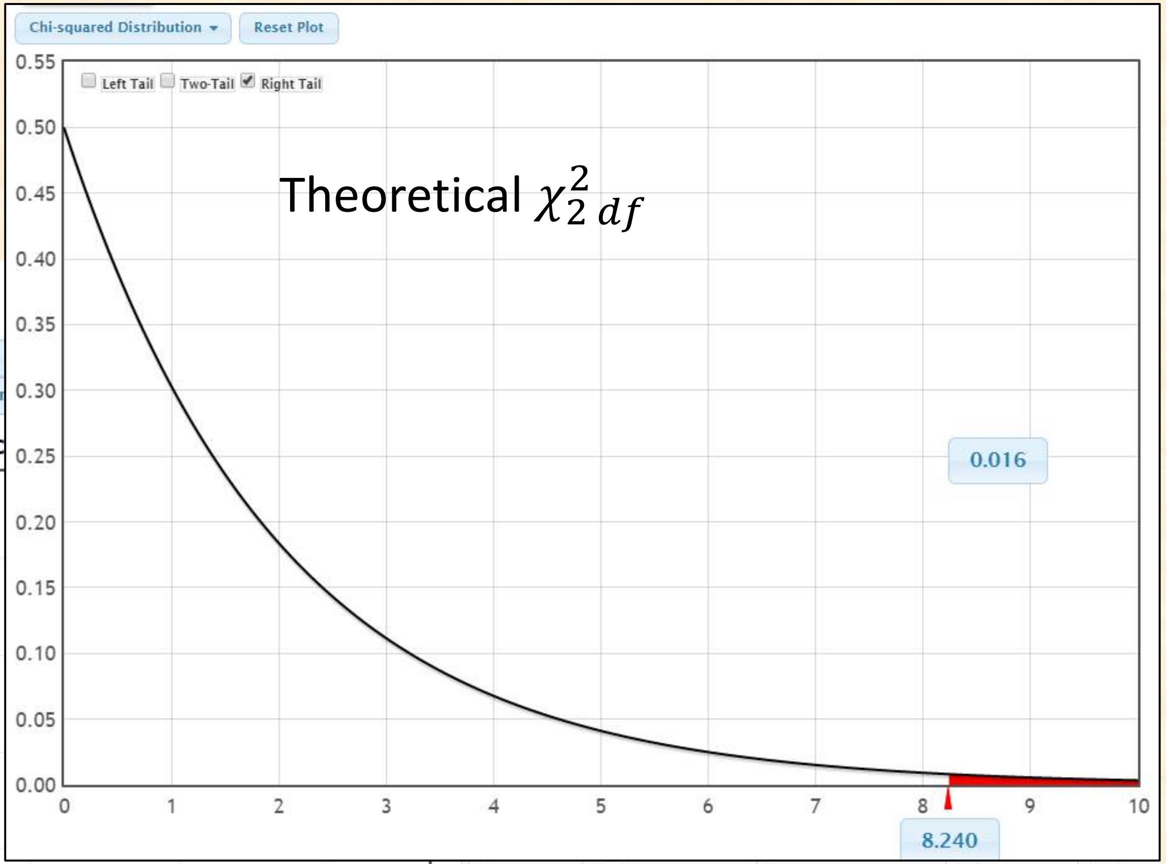
- Shuffle 362 cards (193 male, 169 female)
- Randomly deal 182 cards to Olympic, 31 to Academy, and the remaining 149 to Nobel.
- Find the two-way table (Sex x Award) and compute χ^2.
- Repeat 1,000's of times to get a distribution under the null.

Time for technology…

StatKey

http://lock5stat.com/statkey





F	78	16	75	169
Total	182	31	149	362

Example #2: Sandwich Ants

Experiment:

Place pieces of sandwich on the ground, count how many ants are attracted. Does it depend on filling?

ANOVA Table

	df	SS	MS	F
Groups	2	1561.0	780.5	5.627
Error	21	2913.0	138.7	
Total	23	4474.0		

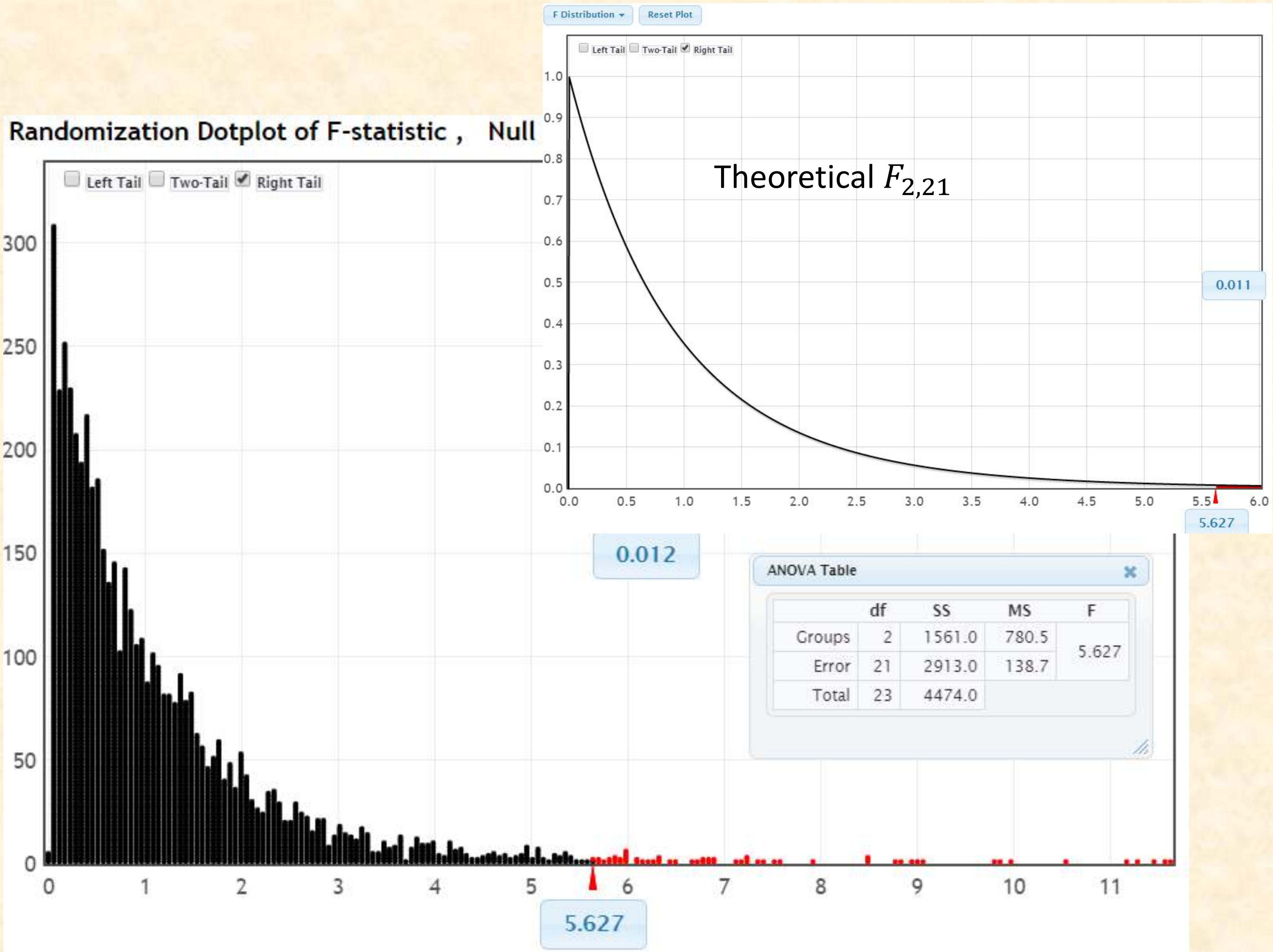
Favourite Experiments: An Addendum to What is the Use of Experiments Conducted by Statistics Students? Margaret Mackisack

http://www.amstat.org/publications/jse/v2n1/mackisack.supp.html

Randomization for Ants

- Write the 24 ant counts on cards.
- Shuffle and deal 8 cards to each sandwich type.
- Construct the ANOVA table and find the F-statistic.
- Repeat 1,000's of times to get a distribution under the null.

StatKey



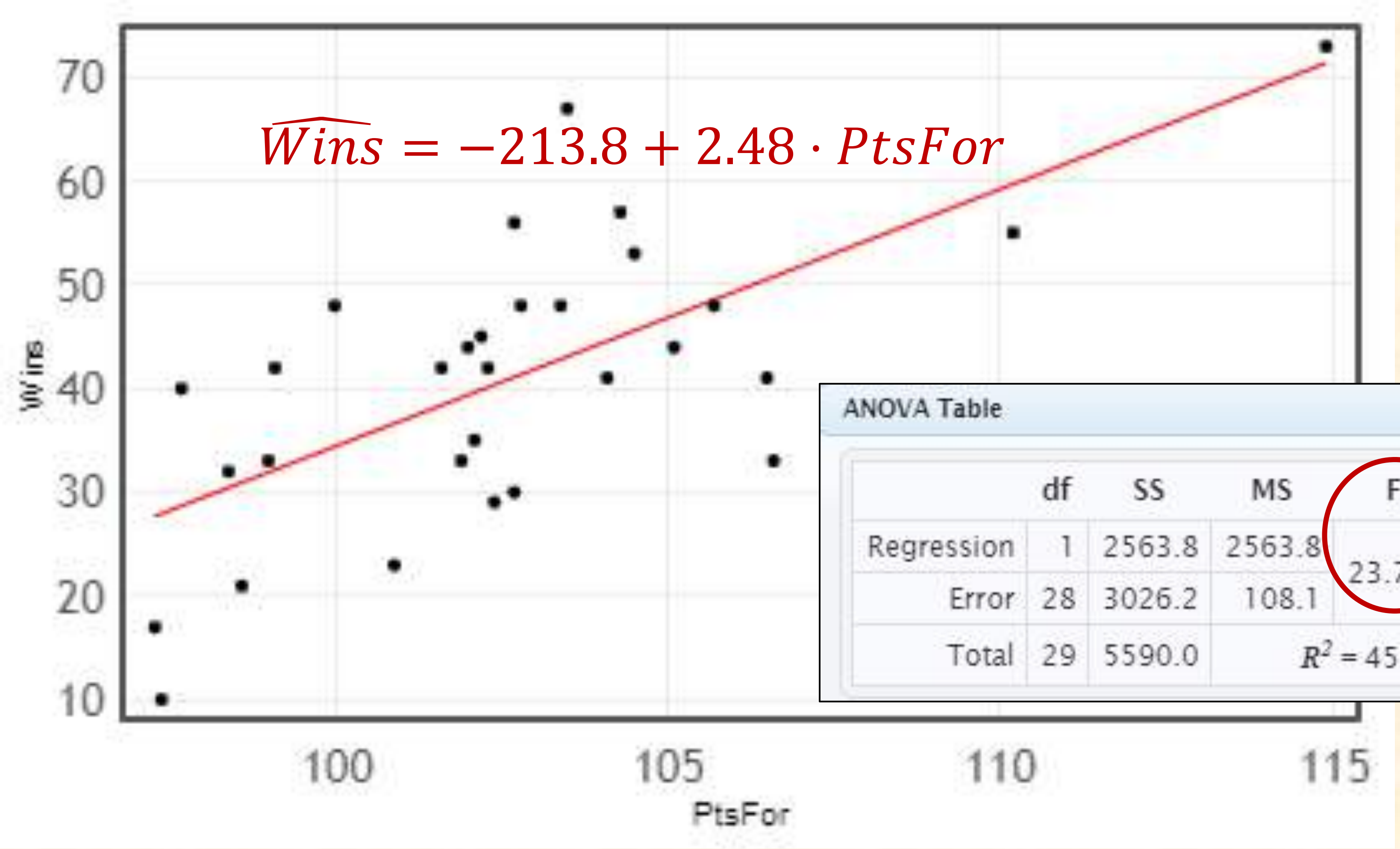
Theoretical $F_{2,21}$

ANOVA Table

	df	SS	MS	F
Groups	2	1561.0	780.5	5.627
Error	21	2913.0	138.7	
Total	23	4474.0		

Example #3: Predicting NBA Wins

Predictor: PtsFor (Points scored per game)



$$\widehat{Wins} = -213.8 + 2.48 \cdot PtsFor$$

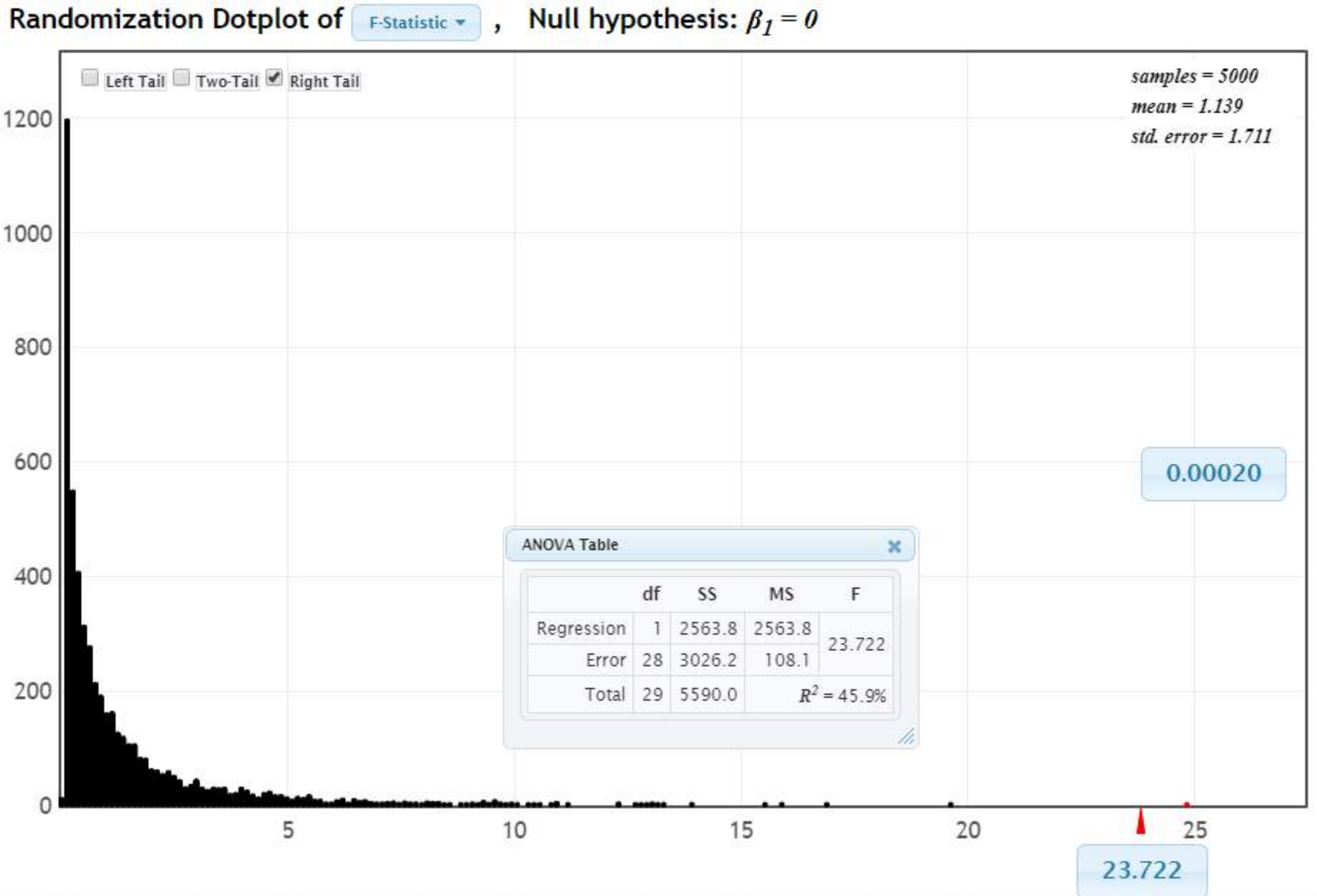
ANOVA Table

	df	SS	MS	F
Regression	1	2563.8	2563.8	23.722
Error	28	3026.2	108.1	
Total	29	5590.0	$R^2 = 45.9\%$	

Randomization for NBA Wins

- Put the 30 win values on cards.
- Shuffle and deal the cards to assign a number of Wins randomly to each team.
- Compute the F-statistic when predicting Wins by PtsFor based on the scrambled sample.
- Repeat 1,000's of times to get a distribution under the null.

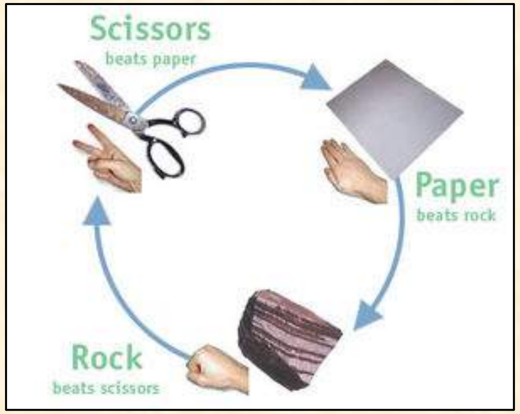
StatKey



	df	SS	MS	F
Regression	1	2563.8	2563.8	23.722
Error	28	3026.2	108.1	
Total	29	5590.0	R^2 = 45.9%	

Example #4: Rock, Paper, Scissors

Play best of three games each.
Record counts for *all* choices.



Rock	Paper	Scissors
65 (72)	67 (72)	84 (72)

n=216

Let p_1, p_2, p_3 be the respective population proportions

$H_0: p_1 = p_2 = p_3 = 1/3$
$H_a: Some\ p_i \neq 1/3$

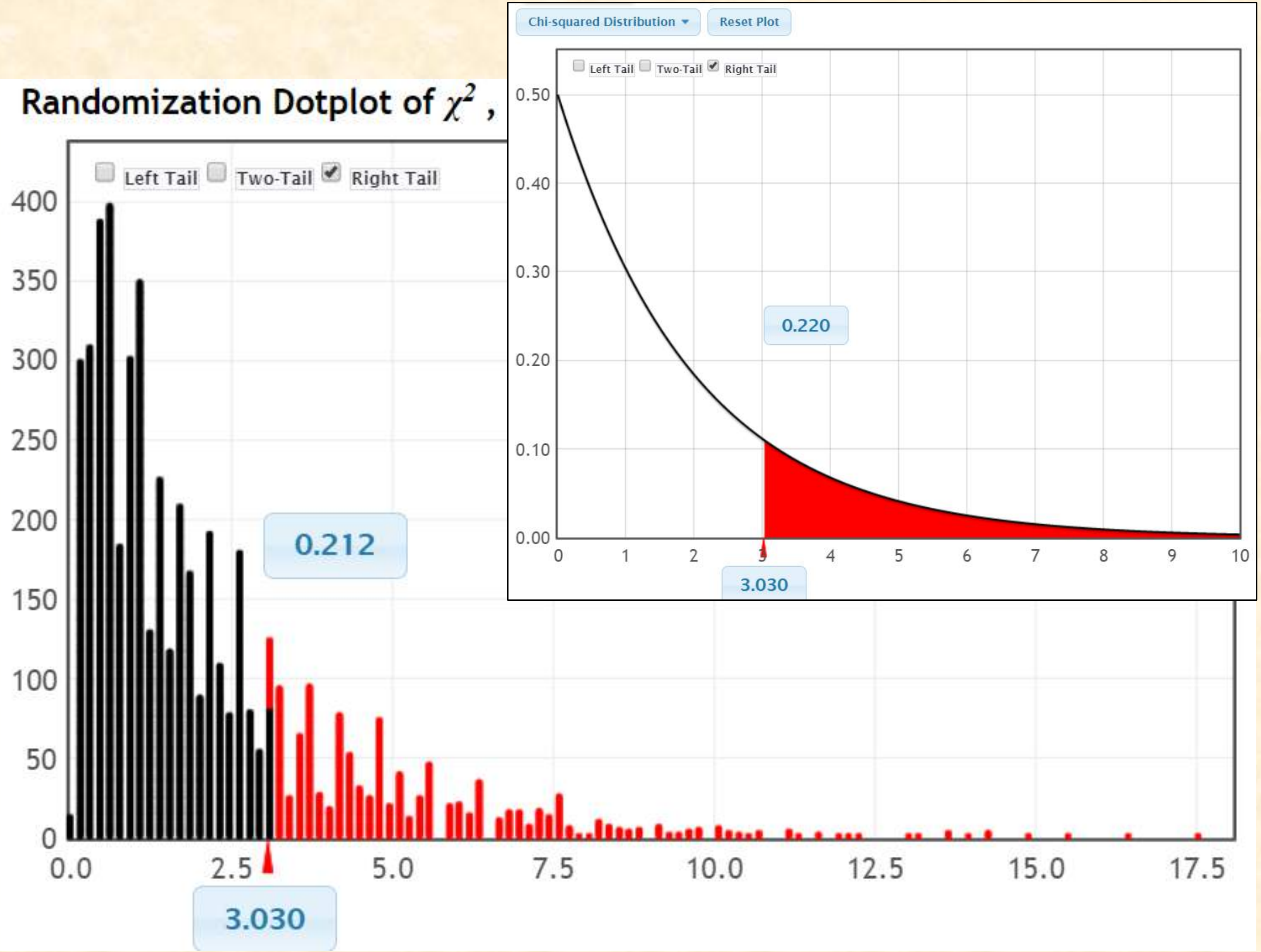
Expected $= np_i = 216 \cdot \frac{1}{3} = 72$

$$\chi^2 = \frac{(65-72)^2}{72} + \frac{(67-72)^2}{72} + \frac{(84-72)^2}{72} = 3.03$$

Randomization for RPS

- Start with an equal number of Rock, Paper, and Scissor cards.
- Sample 216 times *with replacement*.
- Construct the table of counts and compute a chi-square statistic
- Repeat 1000's of times to get a distribution under $H_0: p_1 = p_2 = p_3$.

StatKey


Randomization Dotplot of χ^2 ,
Left Tail
Two-Tail
Right Tail
0.212
3.030
Chi-squared Distribution
Reset Plot
Left Tail
Two-Tail
Right Tail
0.220
3.030

What Statistic?

Chi-square for Association:

$$\chi^2 = \sum \frac{(observed - expected)^2}{expected}$$

ANOVA for Means:

$$F = \frac{MSG}{MSE} = \frac{\sum n_i(\bar{x}_i - \bar{x})^2/df_1}{\sum (x - \bar{x}_i)^2/df_2}$$

ANOVA for Regression:

$$F = \frac{MSModel}{MSE} = \frac{\sum (\hat{y} - \bar{y})^2/df_1}{\sum (y - \hat{y})^2/df_2}$$

If we were ONLY using randomization, would we still use these?

What Statistic?

But StatKey doesn't do that statistic...

```
library(mosaic)
rand_dist=do(5000)*statistic(randomize(data))
```

```
SSqs=do(5000)*anova(lm(sample(y)~x,data=db))
```

```
library(infer)
rand_dist <- data %>%
  specify(y ~ x) %>%
  hypothesize(null = "independence") %>%
  generate(reps = 10000, type = "permute") %>%
  calculate(stat = STATISTIC)
```

Thank you!

QUESTIONS?

Patti Frazer Lock: plock@stlawu.edu
Robin Lock: rlock@stlawu.edu
Kari Lock Morgan: klm47@psu.edu

Slides posted at www.lock5stat.com