

Question: Do real spiders elicit more anxiety than photos of spiders?

Data:	X	Y
	photo	30
	photo	35
	photo	45
	photo	40
	photo	50
	photo	35
	photo	55
	photo	25
	photo	30
	photo	45
	photo	40
	photo	50
	real	40
	real	35
	real	50
	real	55
	real	65
	real	55
	real	50
	real	35
	real	30
	real	50
	real	60
	real	39

$n_p = 12$   
 $\bar{X}_p = 40$   
 $s_p = 9.293$

**test statistic:**  
 $ts = \bar{X}_r - \bar{X}_p = 7$

$n_r = 12$   
 $\bar{X}_r = 47$   
 $s_r = 11.029$

### Randomization-based test:

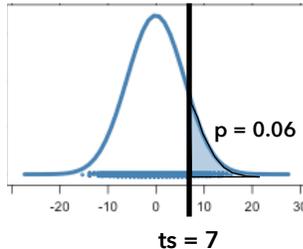
Goal: Estimate the likelihood of obtaining our data under a null hypothesis.

$H_0$ : X has no effect on Y

If we assume  $H_0$  is true, then we can randomly shuffle X with no impact on Y.

- Method:
1. Randomly shuffle X values (keep Y values as they are)
  2. Calculate "randomized" test statistic
  3. Repeat steps 1-2 many times\*
  4. Sketch a plot of the randomized test statistic values
  5. Estimate the likelihood of the actual test statistic (from the data)

Distribution of 10,000 randomized test statistic values (assuming true  $H_0$ )



This p-value was estimated by calculating:

$$\frac{\text{\# of randomized test stats that are } > 7}{10,000}$$

\* The total number of repetitions possible:

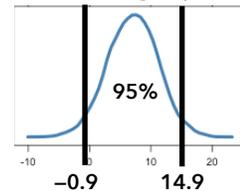
$$\binom{n_1 + n_2}{n_1} / 2$$

### Bootstrap Confidence Interval

Goal: Estimate likely values of the difference between our group means (or whatever our test statistic is measuring).

**Key concept:** Treat our sample data like it's the population. We'll repeatedly sample (with replacement) from our "population" Since we're sampling with replacement, some observations may be chosen multiple times (or never chosen at all)

- Method:
1. Randomly sample **with replacement**  $n_1$  observations from the first group and  $n_2$  observations from the second group
  2. Calculate "bootstrap" test statistic
  3. Repeat many times\*
  4. Sketch a plot of those bootstrap test statistic values
  5. Determine the endpoints of the XX% most likely values of the bootstrap test statistic



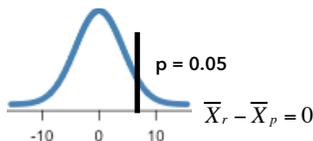
### Theory-based t-test and confidence interval

**t-test goal:** Estimate the likelihood of obtaining our data under a null hypothesis ( $H_0$ : X has no effect on Y)

**CI goal:** Estimate likely values of the difference between our group means (or whatever our test statistic is measuring)

We can sketch the sampling distribution of our test statistic if we assume:

- $H_0$  is true — this allows us to center our sampling distribution at  $\bar{X}_r - \bar{X}_p = 0$
- Our data come from normal populations (or we have large enough sample sizes for the CLT to apply) — this gives us the shape.
- Our treatment groups are independent — this allows us to calculate the fatness (standard error) of our sampling distribution.



**t-test method:**

1. Sketch the sampling distribution (a t-distribution), calculate its standard error.
  - a. Welch-Satterthwaite: Does not assume our treatment groups have equal variances.
  - b. SEpooled: Assumes our treatment groups have equal variances
2. Determine the likelihood of our observed test statistic (using the t-distribution).

**CI method:**

1. Calculate the standard error of our sampling distribution (Welch-S. or SEpooled)
2. Determine the endpoints of the most likely values of the test statistic:  $\bar{X}_r - \bar{X}_p \pm t(SE)$

Effect Size:

$$\text{Cohen's } d = \frac{\bar{X}_r - \bar{X}_p}{s}$$

## R code

- Preliminaries:
1. Make sure to load the mosaic package: `library(mosaic)`
  2. Independent variable **x** and dependent variable **y** are stored in data.frame **d**

### Randomization-based test:

1. Store the test statistic (in this example, the difference between the group means)  
`test.stat <- diffmean(y ~ x, data = d)`
2. Run 10,000 randomizations by shuffling the x variable. Store the randomized test statistics as "rts"  
`rts <- do(10000) * diffmean(y ~ shuffle(x), data = d)`
3. Plot the randomized distribution (as a histogram in this example); calculate and highlight the p-value  
`histogram(~diffmean, data=rts, groups=diffmean >= test.stat,  
main = paste("p = ", prop(~ (diffmean >= test.stat), data=rts)))`

### Bootstrap confidence interval:

1. Store the test statistic (in this example, the difference between the group means)  
`test.stat <- diffmean(y ~ x, data = d)`
2. Generate 10,000 values of the bootstrap test statistic by resampling the data (with replacement). Store the bootstrap values as "boot"  
`boot <- do(10000) * diffmean(y ~ x, data=resample(d))`
3. Plot the bootstrap distribution (as a density plot in this example)  
`densityplot(~diffmean, data=boot, plot.points = FALSE)`
4. Calculate the endpoints of the confidence interval  
`confint(boot, level = 0.95, method = "quantile")`

### t-test:

1. Assuming equal variances and an alternative hypothesis of "less than"  
`t.test(y ~ x, data=d, alternative = c("less"), var.equal = TRUE, conf.level = 0.95)`
2. **Not** assuming equal variances and an alternative hypothesis of "less than"  
`t.test(y ~ x, data=d, alternative = c("less"), var.equal = FALSE, conf.level = 0.95)`

### Theory-based confidence intervals:

1. Assuming equal variances and an alternative hypothesis of "less than"  
`confint(t.test(y ~ x, data=d, conf.level = 0.95, var.equal=TRUE))`
2. **Not** assuming equal variances and an alternative hypothesis of "less than"  
`confint(t.test(y ~ x, data=d, conf.level = 0.95))`