# More About Confidence Intervals and Hypothesis Tests

**Summary:**

- Two ways you could get the wrong answer out of a hypothesis test:
  - **Type I Error**: The null hypothesis is true, but you reject it
  - **Type II Error**: The null hypothesis is incorrect, but you fail to reject it.

- If the null hypothesis is true, the probability of making a Type I Error is $\alpha$, the significance level for the test.
- **Using confidence intervals to conduct hypothesis tests:** If a $(1-\alpha)*100\%$ (example: 95%) CI doesn't contain the value from $H_0$, you can reject $H_0$ at significance level $\alpha$ (example: $\alpha = 0.05$).

### My Favorite Examples of Suspicious Hypothesis Test Results

### Paul the Octopus

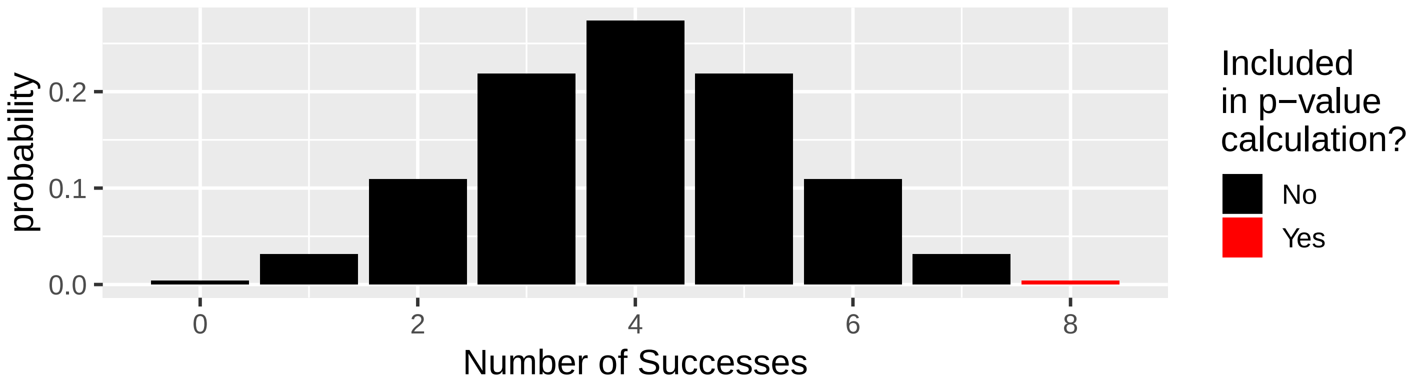
$p$ is the proportion of all predictions Paul might ever make that would be correct.

$H_0$: p = 0.5 (Just guessing!) vs. $H_A$: p > 0.5 (Psychic!!!)

Observed $X = 8$ successes in $n = 8$ trials. If $H_0$ is true, then $X \sim \text{Binomial}(8, 0.5)$ (check conditions). The p-value is:

```
pbinom(q = 7, size = 8, prob = 0.5, lower.tail = FALSE)
```

```
## [1] 0.00390625
```




**Conclusion:** The p-value of 0.0039 is less than $\alpha = 0.01$, so there's very strong evidence that Paul is psychic!!!

### Friday the 13th

$\mu$ is the average difference in number of car accidents on Friday the 13th and on Friday the 6th a week earlier.

$H_0$: $\mu = 0$ (No difference in average number of accidents!)

$H_A$: $\mu > 0$ (Friday the 13th is more dangerous!!!)

In a sample of $n = 6$ pairs, observed:

- sample mean difference of $\bar{y} = 3.333333$
- sample standard deviation of $s = 3.011091$

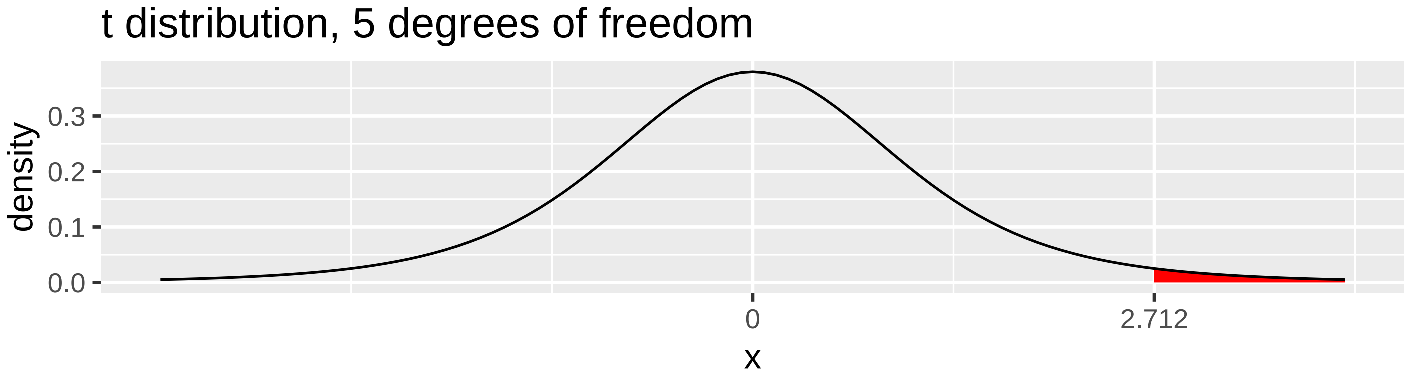
Test statistic: $t = \frac{\bar{y} - \mu_0}{s/\sqrt{n}} = \frac{3.333333 - 0}{3.011091/\sqrt{6}} = 2.712$

If $H_0$ is true, then $t \sim t_5$ (check conditions). The p-value is:

```
pt(2.712, df = 5, lower.tail = FALSE)
```

```
## [1] 0.0210877
```

**Conclusion:** The p-value of 0.021 is less than $\alpha = 0.05$, so there's fairly strong evidence that more accidents occur on Friday the 13th!!!



## Problem 3 from Lab 15: Olympic Swim Suits

In the 2000 Olympics, was the use of a new wetsuit design responsible for an observed increase in swim velocities? In a study designed to investigate this question, twelve competitive swimmers swam 1500 meters at maximal speed, once wearing a new wetsuit and once wearing a regular swimsuit (De Lucas et. al, The effects of wetsuits on physiological and biomechanical indices during swimming. 2000; 3(1): 1-8}). The order of wetsuit versus swimsuit was randomized for each of the 12 swimmers.

```
library(readr)
library(dplyr)
library(ggplot2)

swim <- read_csv("http://www.evanlray.com/data/misc/olympic_wet_suits/olympic
```

```
## Parsed with column specification:
## cols(
##   swimmer_id = col_integer(),
##   wet_suit_velocity = col_double(),
##   swim_suit_velocity = col_double()
## )
```

```
swim <- swim %>%
  mutate(
    velocity_difference = wet_suit_velocity - swim_suit_velocity
  )
```

### (a) What is the population parameter of interest?

**(b) Find a 99% confidence interval for the population mean difference in swim velocity when using the wet suit vs. using the regular swim suit. (All condition checks were confirmed as well as possible in the lab solutions posted on Gryd.)**

```
swim %>%
  summarize(
    mean = mean(velocity_difference),
    sd = sd(velocity_difference)
  )
```

```
## # A tibble: 1 x 2
##     mean          sd
##    <dbl>       <dbl>
## 1 0.0775 0.021794495
```

```
nrow(swim)
```

```
## [1] 12
```

```
qt(0.995, df = 11)
```

```
## [1] 3.105807
```

```
0.0775 - 3.106 * (0.02179449 / sqrt(12))
```

```
## [1] 0.05795852
```

```
0.0775 + 3.106 * (0.02179449 / sqrt(12))
```

```
## [1] 0.09704148
```

**(c) What is the interpretation of the interval found above?**

**(d) Use the confidence interval to conduct the relevant hypothesis test. What is the significance level for the test?**

**(e) What would a Type I error be in this example? Is it possible that a Type I Error was made in part (d)?**

**(f) What would a Type II error be in this example? Is it possible that a Type II Error was made in part (d)?**

## Example: Adapted from SDM4 Part V Wrapup Example 33

A clean air standard requires that vehicle exhaust emissions not exceed specified limits for various pollutants. Many states require that cars be tested annually to be sure they meet these standards. Suppose state regulators double-check a random sample of cars that a suspect repair shop has certified as okay. They will revoke the shop's license if they find significant evidence that the shop is certifying vehicles that do not meet standards.

**(a) Put this in the context of a hypothesis test. What is the population parameter, and what are the hypotheses?**

**(b) What would a Type I error be?**

**(c) What would a Type II error be?**

**(d) Which type of error would the shop's owner consider more serious?**

**(e) Which type of error might environmentalists consider more serious?**

## Example: Adapted from SDM4 Part V Wrapup Example 24

In June 2010, a random poll of 800 working men found that 72 of them had taken on a second job to help pay the bills (www.careerbuilder.com). Let's assume that this sample was representative of working men. Based on these data, a 95% confidence interval for the proportion of working men who had taken on a second job is [0.071, 0.112].

**(a) A pundit on a TV news show claimed that 6% of working men had a second job. Use the confidence interval above to test whether this claim is plausible given the poll data. What is the significance level of your test?**

**(b) Is it possible that the pundit was correct, and you made an error in your hypothesis test from part (a)? If so, what type of error would that be?**

**(c) For what proportion of samples would the testing procedure you used in part a result in a Type I error, if the null hypothesis was true?**

**(d) For what proportion of samples would the confidence interval from the problem statement contain the true proportion of working men who had taken on a second job?**