# 20: TWO SAMPLE PROPORTION TESTS + CONNECTION TO CONFIDENCE INTERVALS

Stat250 S25

Prof Amanda Luby

---

**Example: A/B Testing** Online controlled experiments are a digital version of randomized controlled trials, where the designer randomly assigns participants to see a different version of a website. These are colloquially referred to as "A/B" tests. For example, Google famously tested "50 shades of blue" when determining which color to use for links in ads, where they showed users the same webpage, where the only difference was the color of blue used in a link. If more users clicked on the link for one shade of blue than another, they used that shade of blue. This experiment reportedly led to a $200M increase in ad revenue (see *Why Google has 200m reasons to put engineers over designers* in **The Guardian**)

Let's pretend we run an A/B test on 200 users, showing links in Cornflower Blue or Midnight Blue. Since the stakes of making a mistake are not too high, we'll use $\alpha = .15$.

| | Clicked | Did Not Click |
|---|---|---|
| Cornflower Blue | 26 | 74 |
| Midnight Blue | 22 | 78 |

## 1 Two-sample proportion tests

**Sample Statistic:**

**Test Statistic:**

**Sampling distribution under $H_0$:**

```
prop.test(x = c(26, 22), n = c(26+74, 22+78), alternative = "two.sided", conf.level = .85)
```

```
    2-sample test for equality of proportions with continuity correction

data:  c out of c26 out of 26 + 7422 out of 22 + 78
X-squared = 0.24671, df = 1, p-value = 0.6194
alternative hypothesis: two.sided
85 percent confidence interval:
 -0.05685041  0.13685041
sample estimates:
prop 1 prop 2
  0.26   0.22
```

## 1.1 Matched Pairs for Proportions

**Example**: Oops! These are the same 100 people on two different visits, with the order of the color assignment randomized.

| | Clicked | Did Not Click |
|---|---|---|
| Clicked | 12 | 10 |
| Did Not click | 14 | 64 |

**McNemar's test for paired proportions**

```
prop.test(10, 24, p = .5, conf.level = .85)
```

```
    1-sample proportions test with continuity correction

data:  10 out of 24
X-squared = 0.375, df = 1, p-value = 0.5403
alternative hypothesis: true p is not equal to 0.5
85 percent confidence interval:
 0.2659953 0.5824121
sample estimates:
        p
0.4166667
```

## 2 Connection between Confidence Intervals and Hypothesis Tests

### Inverting a hypothesis test

```
binom.test(10, 24, p = .5, conf.level = .85)
```

```
data:  10 out of 24
number of successes = 10, number of trials = 24, p-value = 0.5413
alternative hypothesis: true probability of success is not equal to 0.5
85 percent confidence interval:
 0.2636000 0.5830746
sample estimates:
probability of success
             0.4166667
```