Economics 103 – Statistics for Economists

Francis J. DiTraglia

University of Pennsylvania

Optional Addendum to Lecture 23

Example of Calculating Power: Is this coin fair?

Flip a Possibly Unfair Coin n Times

$X_1, \ldots, X_n \sim$ iid Bernoulli(p) where p may not equal $1/2$

Test $H_0: p = 0.5$ against $H_1: p \neq 0.5$

Under H_0 and if n is large the CLT gives

$$T_n = \frac{\widehat{p} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{n}}} \approx N(0,1)$$

The Idea Behind Statistical Power

How does the test statistic behave if the H_0 is *false*? What is the sampling distribution of T_n *under the alternative*: $H_1: p \neq 0.5$

Key Distributional Result

Center and standardize $\widehat{p}$ using *true* $p \implies$ standard normal:

$$Z = \frac{\widehat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \approx N(0,1)$$

Under the Alternative $p \neq 0.5$

T_n is *incorrectly centered and scaled*: $T_n = \dfrac{\widehat{p} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{n}}} \neq N(0,1)$

How We'll Proceed

Use algebra to express T_n in terms of Z, which we know is $N(0,1)$, and constants. This will give the distribution of T_n under H_1.

This is Just Algebra

$$
\begin{aligned}
T_n &= \frac{\widehat{p} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{n}}} = \sqrt{n}(2\widehat{p} - 1) \\
&= \sqrt{n}\left[2\widehat{p} - 1 + (2p - 2p)\right] = \sqrt{n}\left[2\left(\widehat{p} - p\right) + 2p - 1\right] \\
&= \sqrt{n}\left[2\left(\widehat{p} - p\right)\left(\frac{\sqrt{p(1-p)/n}}{\sqrt{p(1-p)/n}}\right) + 2p - 1\right] \\
&= \sqrt{n}\left[2\sqrt{\frac{p(1-p)}{n}}\left(\frac{\widehat{p} - p}{\sqrt{p(1-p)/n}}\right) + 2p - 1\right] \\
&= \left[2\sqrt{p(1-p)}\right] Z + \sqrt{n}(2p - 1) \\
&= a\, Z + b
\end{aligned}
$$

Distribution of Test Statistic Under H_1: $p \neq 0.5$

From the previous slide,

$$T_n = \frac{\widehat{p} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{n}}} = aZ + b$$

where:

$$\begin{aligned} Z &= \frac{\widehat{p} - p}{\sqrt{p(1-p)/n}} \approx N(0,1) \\ a &= 2\sqrt{p(1-p)} \\ b &= \sqrt{n}(2p-1) \end{aligned}$$

Hence: $T_n \approx N(\mu = b, \sigma^2 = a^2)$

Distribution of T_n Under the Alternative

$$\boxed{T_n = \frac{\widehat{p} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{n}}} \approx N\left(\sqrt{n}(2p-1), 4p(1-p)\right)}$$

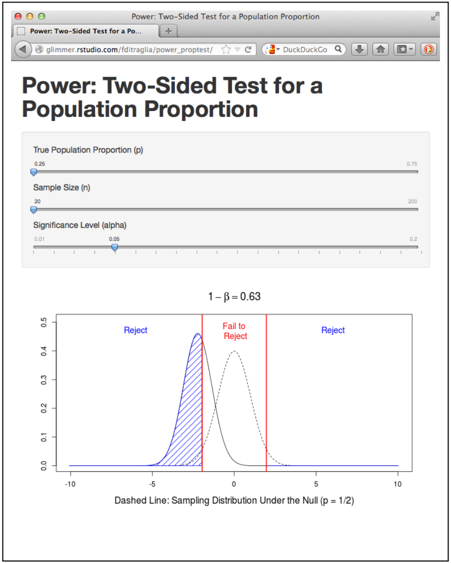
Note That:

1. Mean depends on p and n
2. Variance depends only on p
3. If $p = 0.5$ so H_0 is true, this reduces to a standard normal:

$$\begin{aligned} \text{Mean} &= \sqrt{n}(2p-1) = \sqrt{n}(2 \times 1/2 - 1) = 0 \\ \text{Variance} &= 4p(1-p) = 4 \times 1/2 \times (1 - 1/2) = 1 \end{aligned}$$

http://glimmer.rstudio.com/fditraglia/power_proptest/

Ignore everything except the solid curve and play around with the first two sliders.



Now We can Calculate Power!

Decision Rule for Two-sided Test

Reject $H_0: p = 0.5$ provided that $|T_n| \geq \texttt{qnorm}(1-\alpha/2)$

If the null is false (i.e. under $H_1: p \neq 0.5$)

$$T_n = \frac{\widehat{p} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{n}}} \approx N\left(\sqrt{n}(2p-1), 4p(1-p)\right)$$

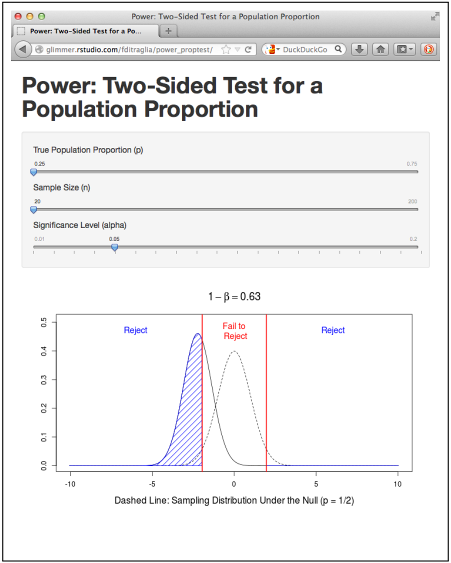
Thus, the probability of rejecting a false null is:

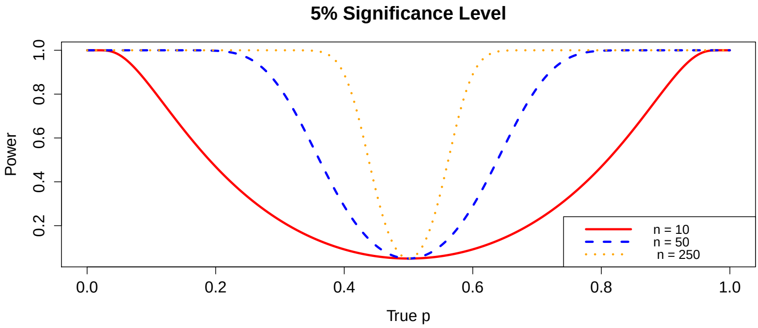
$$\boxed{\text{Power}(\alpha, p, n) = P\left(|Y| \geq c\right) = P(Y \leq -c) + P(Y \geq c)}$$

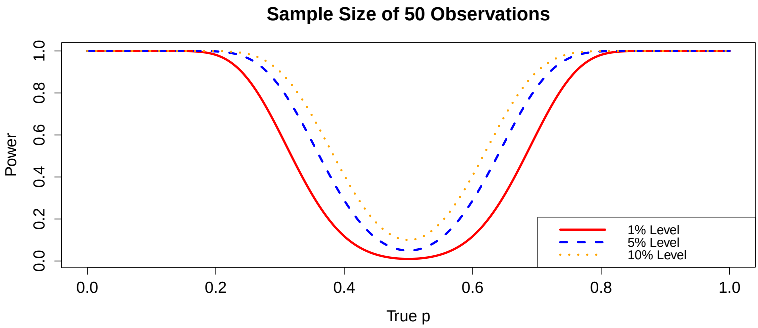
$$\begin{aligned} c &= \texttt{qnorm}(1-\alpha/2) \\ Y &\sim N\left(\sqrt{n}(2p-1), 4p(1-p)\right) \end{aligned}$$

http://glimmer.rstudio.com/fditraglia/power_proptest/

Now look at everything and try changing all the sliders!




5% Significance Level
Power
True p
0.0
0.2
0.4
0.6
0.8
1.0
n = 10
n = 50
n = 250


Sample Size of 50 Observations
Power
True p
0.0
0.2
0.4
0.6
0.8
1.0
1% Level
5% Level
10% Level

Some Intuition about Power for the Coin Example

- Equals prob. of rejecting false null, i.e. convicting a guilty person.
- Tells us how large a sample we would need to detect a given discrepancy from "fairness" of the coin: $p = 0.5$
  - Small deviations from $p = 0.5$ unlikely to be detected unless the sample size is large.
  - Large deviations from $p = 0.5$ very likely to be detected even if the sample size is small.
- For a *given* degree of "unfairness" (deviation from $p = 0.5$)
  - Higher significance level (α) $\implies$ higher power $(1 - \beta)$
  - Large sample size (n) $\implies$ higher power $(1 - \beta)$

Power More Generally

Power $= (1 - \beta) = 1 - P(\text{Type II Error})$

Chance of detecting an effect given that one exists.

Power Depends on Four Things

1. Magnitude of Effect: easier to detect large deviations from H_0
2. Amount of variability in the population: less variability $\implies$ easier to detect an effect of given size.
3. Sample Size: larger $n \implies$ easier to detect effect of given size
4. Signif. Level (α): fewer Type I errors $\implies$ more Type II errors

Go back and compare to factors that affect width of CI...

An Important Point

For any discrepancy from $p = 0.5$, there is always a sample size large enough to make the power of our test *arbitrarily close to one*. In other words, we can be almost certain to detect an effect, no matter how small, provided that we use a large enough sample. However, this does not mean that the affect we have found is important.

If it turned out that the true probability of heads was closer to 0.50001 rather than 0.5, should we really care?

An R Function to Calculate Power For Coin Example

$$
\begin{aligned}
\text{Power}(\alpha, p, n) &= P\left(Y \leq -c\right) + P\left(Y \geq c\right) \\
c &= \texttt{qnorm}(1 - \alpha/2) \\
Y &\sim N\left(\sqrt{n}(2p - 1), 4p(1 - p)\right)
\end{aligned}
$$

```
coin.power <- function(a, p, n){
    c <-  qnorm(1 - a/2)
    mu <- sqrt(n) * (2 * p - 1)
    sigma <- sqrt(4 * p * (1 - p))

    less.than <- pnorm(-c, mean = mu, sd = sigma)
    greater.than <- 1 - pnorm(c, mean = mu, sd = sigma)
    power <- less.than + greater.than
    return(power)
}
```

Here's Some Code for You to Play Around With

You can use the function `coin.power` to calculate power for specific values of p, n, and α

```
coin.power(a = 0.05, p = 0.55, n = 100)
coin.power(a = 0.05, p = 0.55, n = 1000)
coin.power(a = 0.1, p = 0.55, n = 100)
coin.power(a = 0.05, p = 0.6, n = 100)
```

or to make plots similar to those on slide 28

```
alternatives <- seq(from = 0, to = 1, by = 0.001)
power <- coin.power(a = 0.05, alternatives, n = 10)
plot(alternatives, power, xlab = 'True p', ylab = 'Power', type = 'l')
```

Try out different choices for p, n and α and see what you get! I'll post this R code to save typing.