Advanced Econometrics

University of Warsaw

Jerzy Mycielski

Doctoral School of Social Sciences, 2024

Significance test (t - test)

- Consider null H_0 and alternative hypotheses H_1

$$H_0 : \beta_j = \beta$$

$$H_1 : \beta_j \neq \beta$$

- We reject H_0 (and accept H_1) if $|T| > c$ as large value of $|T|$ suggests that $\widehat{\beta}$ deviates *significanly* from β.
- We can make two errors
  - reject H_0 which is true (type 1 error)
  - not reject H_0 which is false (type 2 error)
- Probability of type 1 error is equal to

$$\mathbb{P}\left[|T| > c | H_0 \text{ true} \right] = \alpha$$

  where α is called significance level
- Probability of type 2 error

$$\mathbb{P}\left[|T| < c | H_0 \text{ false} \right] = \beta$$

  $1 - \beta$ (probability of rejecting false H_0) is known as the power of the test

Size and power, p-value

- When testing hypothesis we set certain significance level.
- The conventional significance level used in economics is $\alpha = 0.05$ and for power $1 - \beta = 0.8$.
- The power of the test can usually be calculated only for some assumed value of β. The smaller is the effect size β the smaller is the power of the test.
- Estimation precision increases with the number of observations, sample size should be sufficient to provide adequate power for testing the assumed effect size.
- Notice that if distribution of T is symmetric

$$\mathbb{P}\left[\left|T\right| > c \middle| H_0 \text{ true}\right] = 2\left(1 - F\left(c\right)\right) = \alpha$$

  and $2\left(1 - F\left(c\right)\right)$ is decreasing function of c.
- Therefore for same realized value of the statistic t

$$|t| > c \iff 2\left(1 - F\left(|t|\right)\right) < \alpha$$

- Instead of comparing $|t|$ with critical value c we can compare $2\left(1 - F\left(|t|\right)\right)$ (called p-value) with α.

```
      Source |       SS           df       MS      Number of obs   =        74
-------------+----------------------------------   F(2, 71)        =     69.75
       Model |   1619.2877         2  809.643849   Prob > F        =    0.0000
    Residual |  824.171761        71   11.608053   R-squared       =    0.6627
-------------+----------------------------------   Adj R-squared   =    0.6532
       Total |  2443.45946        73  33.4720474   Root MSE        =    3.4071
------------------------------------------------------------------------------

         mpg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      weight |  -.0065879   .0006371   -10.34   0.000    -.0078583   -.0053175
     foreign |  -1.650029   1.075994    -1.53   0.130      -3.7955    .4954422
       _cons |    41.6797   2.165547    19.25   0.000     37.36172    45.99768
```

Source: *Stata 18 Base Reference Manual.* (2023).

Replication crises

- In majority of cases in economics and in other sciences researchers are interested in cases when $\beta_j \neq 0$ - variable x_j influences the outcome variable y
- In such a case we are testing significance of of variable β_j by testing $H_0 : \beta_j = 0$ under $H_1 : \beta_j \neq 0$
- Obviously the desired result is the rejection of H_0 and acceptance of H_1 of statistical significance of x_j
- It is now well known that large proportion of studies (not only in economics) which have shown some significant relationships failed to be replicated
- By replication we mean repeating the same study on different data set
- Why this is the case?

Positive predictive value

- John P A Ioannidis (2005) formulated simple formula which can help explain why large proportion of research findings are false.
- Assume that some proportion of tested hypotheses H_1 are valid. This proportion can be interpreted as unconditional probability that H_1 is true.
- If research hypothesis is credible than this probability is high.
- Denote $(H_1 : \mathit{true})$ as T, $(H_1 : \mathit{false})$ as F
- Positive result of the test (H_1 accepted) is denoted as P,
- Probability that H_1 is true given that H_1 was accepted (Positive Predictive Value) is equal to:

$$\mathbb{P}(T|P) = \frac{\mathbb{P}(T \wedge P)}{\mathbb{P}(P)} = \frac{\mathbb{P}(P|T)\,\mathbb{P}[T]}{\mathbb{P}[P|T]\,\mathbb{P}[T] + \mathbb{P}[P|F]\,[\mathbb{P}(F)]}$$

$$= \frac{(1-\beta)\,\mathbb{P}(T)}{(1-\beta)\,\mathbb{P}[T] + \alpha\,(1-\mathbb{P}[T])}$$

- The positive predictive value can be interpreted as the percentage of research findings which are indeed true.
- Notice that it depends on α but also on $\mathbb{P}(P)$ and β
- Example:
  - $\mathbb{P}(T) = \frac{1}{2}$ (50% hypothesis tested are valid), conventional $\beta = 0.2$ (power 80%), $\alpha = 0.05$: $PPV = 0.94$
  - $\mathbb{P}(T) = 0.3$ (30% hypothesis tested are valid), underpowered $\beta = 0.8$ (power 20%), $\alpha = 0.05$: $PPV = 0.63$
- John P A Ioannidis (2005) suggested as well that some research findings are due to undetected mistakes in the research procedure. If the proportion of such findings is u then

$$\mathbb{P}(T|P) = \frac{(1-(1-u)\beta)\mathbb{P}[T]}{(1-(1-u)\beta)\mathbb{P}[P] + ((1-u)\alpha + u)(1-\mathbb{P}[T])}$$

- Example:
  - $\mathbb{P}(T) = 0.3$ (30% hypothesis tested are valid), underpowered $\beta = 0.8$ (power 20%), $\alpha = 0.05$, $u = 0.1$: $PPV = 0.45$

Meta-analysis of power in economic research

- The main problem with measuring power is that the β effect size is unknown
- However, if the multiple studies were published estimating the same β then it can be estimated as the (weighted) average $\overline{\beta}$ of the estimates published
- The power of individual studies can then be estimated by comparing the estimated standard errors with the estimate $\overline{\beta}$ of the effect β
- Such meta-analysis was done by John P. A. Ioannidis, Stanley, and Doucouliagos (2017) on the basis of 159 meta-analyses, 6730 primary articles and 64076 empirical estimates.
- Results suggest that in most of the research areas in economic the empirical studies are seriously underpowered and the estimates of the parameters are inflated

Estimated powers for articles in economics

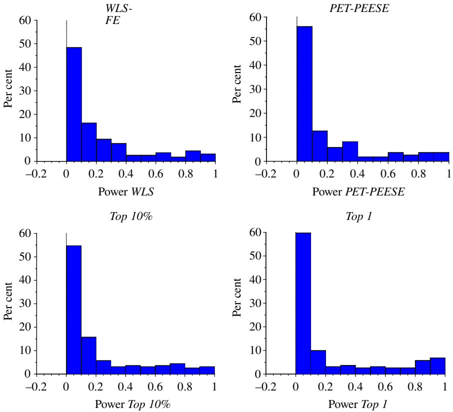

Fig. 1. Histograms of Adequately Powered Estimates in Empirical Economics
Note. Colour figure can be viewed at wileyonlinelibrary.com.

Source: John P. A. Ioannidis, Stanley, and Doucouliagos (2017)

Publication bias and p-hacking

- Journals seems to prefere articles with significant results - this is known as publication bias
- This creates adverse incentive for researchers to modify (possibly unconsciously) the methodology of the study so to obtain significant results
- The flexibility of definitions, the choice of estimation methods, the possible transformations of the data often make this practice possible
- It is called as p-hacking

Publication bias and p-hacking

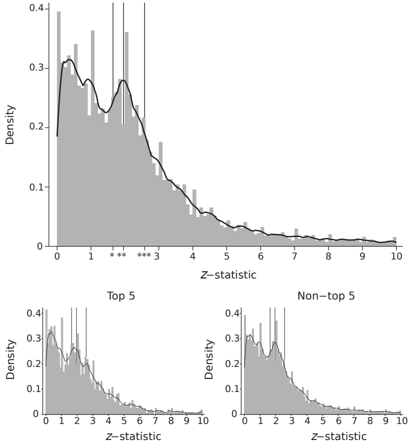

Figure 1. *z*-Statistics in 25 Top Economics Journals

Notes: The top panel displays histograms of test statistics for $z \in [0, 10]$. Bins are 0.1 wide. Reference lines are displayed at the conventional two-tailed significance levels. We have also superimposed an Epanechnikov kernel. The bottom left panel presents test statistics from the Top 5 journals (*American Economic Review, Econometrica, Journal of Political Economy, Quarterly Journal of Economics, and Review of Economic Studies*). The bottom right panel presents test statistics from the remainder of the sample. We do not weight articles.

Source: Brodeur, Cook, and Heyes (2020)

What determines the credibility of the studies

- John P A Ioannidis (2005) suggested that research findings are less likely to be true:
  - the smaller the sample used.
  - the smaller are the estimated effect β.
  - the greater the number and the lesser the selection of tested hypotheses.
  - the greater the flexibility in designs, definitions, outcomes, and analytical modes.
  - the greater the financial and other interests and prejudices.
  - the hotter the scientific field (with more scientific teams involved).

Bibliography

- Brodeur, Abel, Nikolai Cook, and Anthony Heyes (Nov. 2020). "Methods Matter: p-Hacking and Publication Bias in Causal Analysis in Economics". In: *American Economic Review* 110.11, pp. 3634–60. DOI: 10.1257/aer.20190687. URL: https://www.aeaweb.org/articles?id=10.1257/aer.20190687.
- Ioannidis, John P A (Aug. 2005). "Why Most Published Research Findings Are False". In: *PLOS Medicine* 2.8, pp. 1–1. DOI: 10.1371/journal.pmed.0020. URL: https://ideas.repec.org/a/plo/pmed00/0020124.html.
- Ioannidis, John P. A., T. D. Stanley, and Hristos Doucouliagos (2017). "The Power of Bias in Economics Research". In: *The Economic Journal* 127.605, F236–F265. DOI: https://doi.org/10.1111/ecoj.12461. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ecoj.12461. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/ecoj.12461.
- *Stata 18 Base Reference Manual.* (2023).