CUTR
CENTER for URBAN TRANSPORTATION RESEARCH

Testing Equivalence of Measurements & Application to Certifying APC Data for NTD Reporting

Xuehao Chu

March 17, 2016 • CUTR Webcast Series

Outline

- Motivation
- Testing Equivalence in General
- Application to Certifying APC Data for NTD Reporting
- Comments on Federal Register Notice
- Contact Info

National Transit Database (NTD)

- U.S. public transit agencies that benefit from Federal transit grant programs report their data to the NTD on:
  - Resource used
  - Service provided
  - Service consumed
- Service consumed includes:
  - Unlinked passenger trips (UPT), i.e., boardings
  - Passenger miles traveled (PMT)
- Service consumed: 100% count or estimated
  - Annual UPT: increasingly a 100% count
  - Annual PMT typically estimated
- **Minimum 10% precision & 95% confidence**

Motivation 3

Load-Based PMT for Each Trip

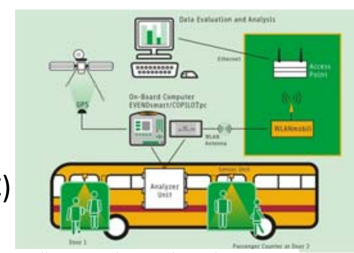
1. Collect detailed field data:
   - Boardings at each stop
   - Alightings at each stop
2. Balance trip-total boardings and alightings
3. Calculate loads between stops
4. Calculate PMT between stops
5. Sum PMT between stops across all stops

Motivation 4

Manual Method for Field Data

- Human ride checks or in-vehicle video
- Advantages
  - Can ALSO observe the following
    - Passengers from the previous trip at 1st stop
    - Passenger loading between each pair of stops
    - Passengers continuing to the next trip at last stop
  - Minimum measurement error when done properly
- Disadvantages
  - High labor cost
  - Limited coverage (i.e., a small sample of trips each year)
- Estimate annual PMT through random sampling

Automated Method



http://www.init-ka.de/gfx_content/products/APC_55_large.jpg

- Automatic passenger counter (APC)
- Advantages
  - Low labor cost
  - Wide coverage (i.e., all trips with APC)
- Disadvantages
  - Cannot also observe the following
    - Passengers from the previous trip at 1st stop
    - Passenger loading between each pair of stops
    - Passengers continuing to the next trip at last stop
  - Potentially high measurement error in PMT
  - Worst with interlined and loop services
  - Wide variation in accuracy across applications
- Estimate annual PMT through
  - Using all valid APC data, or
  - Using random sampling

Components of Errors in Annual PMT

- Manual Data
  - Zero measurement error (at least assumed so)
  - Pure random error from sampling
  - Sampling plans are based on 10% precision
- APC Data
  - Using all valid APC data
    - Minor random error due to large sample size
    - Mainly measurement error
  - Using random sampling
    - Large random error due to small sample size
    - Measurement error can be significant
    - Sampling plans must account for measurement error: sampling precision = 10% - measurement error

Certification of APC Data

- NTD requires APC data be certified for when used for reporting
- NTD requires proof of equivalence, i.e., APC and manual data differ within an acceptable difference
- NTD requires using paired observations, i.e., collecting both APC and manual data from the same sample of vehicle trips

Potential Methods for Certification

- Non-Statistical Testing
  - Test if APC and manual data differ within an acceptable difference **numerically**
  - Referred to as *Numerical Testing* here
- Statistical Testing
  - Traditional t-test:
    - referred to as *Difference Testing*
  - A well established test in medicine
    - test if APC and manual data differ within an acceptable difference **statistically**
    - referred to as *Equivalence Testing*

Testing Equivalence 9

Basics of Statistical Testing

- A pair of hypotheses:

  Null (H_0) vs Alternative (H_1)
- Objective is to prove H_1 through rejecting H_0
- Not rejecting H_0 means not enough evidence to reject it
- Not rejecting H_0 does not prove it statistically
- H_1 must be based on what one wants to prove:
  - To detect difference requires equality in H_0 and inequality in H_1 (e.g., traditional t-test)
  - To prove "equivalence" requires "non equivalence" in H_0 and "equivalence" in H_1

Testing Equivalence 10

Basic Terms and Notations

- *n* paired observations: y_{ai} (automated) and y_{mi} (manual)
- Mean for manual data: $\overline{y_a} = \frac{\sum y_{ai}}{n}$
- Difference for each pair: $d_i = y_{ai} - y_{mi}$
- Mean difference: $\bar{d} = \frac{\sum d_i}{n}$ (or $\frac{\bar{d}}{\overline{y_a}}$ in percent terms)
- Standard deviation of difference: $s_d^2 = \frac{\sum(d_i - \bar{d})^2}{n-1}$
- Standard error of mean difference: $s_{\bar{d}} = \frac{s_d}{\sqrt{n}}$
- Significance level for statistical testing: α (e.g., 5%)
- Population mean difference: μ_d (the true value)
- Acceptance criterion for *Numerical Testing*: δ
- Acceptance criterion for *Equivalence Testing*: θ

CUTR

Testing Equivalence

11

Numerical Testing

- Pre-select a specific value for δ (i.e., acceptance criterion)
- Compare $\bar{d}$ to δ numerically:
  - Accept equivalence of APC data if $|\bar{d}| \leq \delta$
- Simple
- Ignore variability in $\bar{d}$ across samples
- Consider precision but not significance
- High probability of accepting equivalence by chance
- Do not ensure 10% precision at 95% confidence

CUTR

Testing Equivalence

12

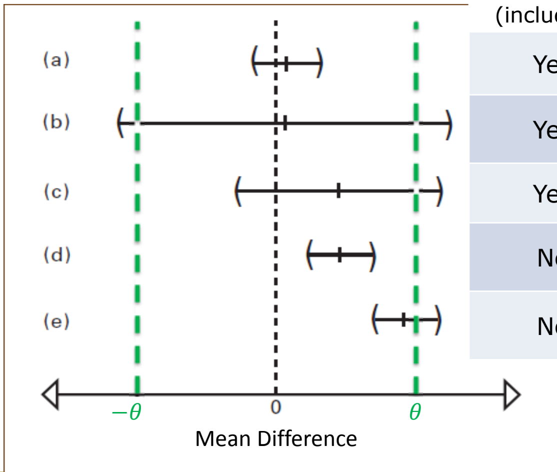
Difference Testing

- Hypotheses: H_0: $\underbrace{\mu_d = 0}_{\text{Equal}}$; H_1: $\underbrace{\mu_d \neq 0}_{\text{Not Equal}}$
- Designed to detect difference through rejecting equality
- Failing to reject H_0 if
  - $t = \frac{|\bar{d}|}{s_{\bar{d}}} \leq z_{1-\alpha/2}$
  - Or $(1\text{-}\alpha)$% confidence interval includes 0
- Failing to reject equality is mistaken with proving equality (equivalence)
- Poor precision or small sample favors this mistaken equivalence

Equivalence Testing

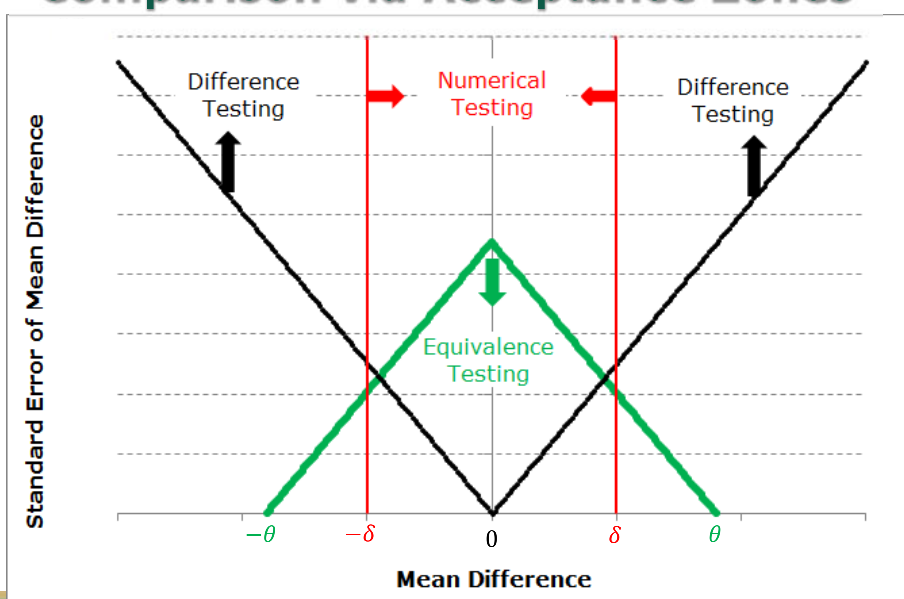
- Pre-select acceptance criterion θ:
  - mean difference not exceeding θ is acceptable
  - Based on prior knowledge and application at hand
- Hypotheses: H_0: $\underbrace{\mu_d \leq -\theta \text{ or } \mu_d \geq \theta}_{\text{Not Equivalent}}$; H_1: $\underbrace{|\mu_d| < \theta}_{\text{Equivalent}}$
- Designed to prove equivalence through rejecting not being equivalent
- Use well established Two One-Sided Test (TOST)
- Reject H_0 (i.e., accept equivalence at α) if
  - $t_1 = \frac{\bar{d}-(-\theta)}{s_{\bar{d}}} > z_{1-\alpha}$ and $t_2 = \frac{\bar{d}-\theta}{s_{\bar{d}}} < -z_{1-\alpha}$
  - Or $(1\text{-}2\alpha)$% confidence interval falls within ($-\theta$, θ)
- Avoid issues with Numerical or Difference Testing

Comparison via Confidence Intervals

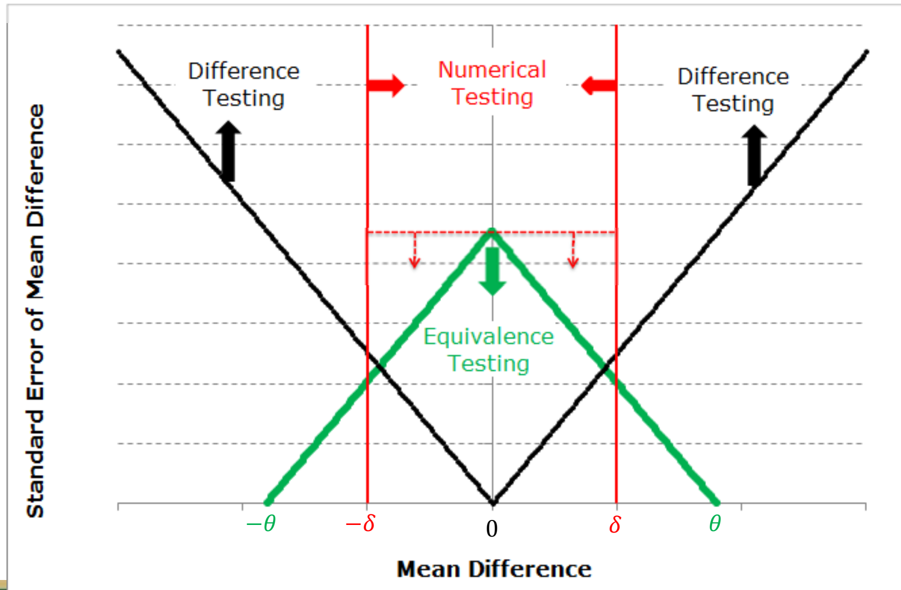


Equivalent?	
Difference Testing (include 0)	***Equivalence Testing*** $(-\theta, \theta)$
Yes	Yes
Yes	No
Yes	No
No	Yes
No	No

Comparison via Acceptance Zones



Cap Standard Error for *Numerical Testing*



Testing Equivalence

17

Pre-Selection of Acceptance Criterion for *Equivalence Testing*

- Highly encourage using all valid APC data:
  - A reasonable value for $\theta = 9\%$, allowing a small amount of up to 1% random error due to <100% APC coverage and invalid APC data.
- If using random sampling for estimation:
  - Pre-select θ = 6%, leaving adequate room for minimum sampling precision
  - More difficult for APC data to be certified
  - Minimum sampling precision = 10% - actual $|\bar{d}|$
    - If actual $|\bar{d}|$ = 5%, sampling precision must be increased to 5%. As a result, annual sample size would **quadruple** over what 10% sampling precision would require.

Application to Certifying APC Data

18

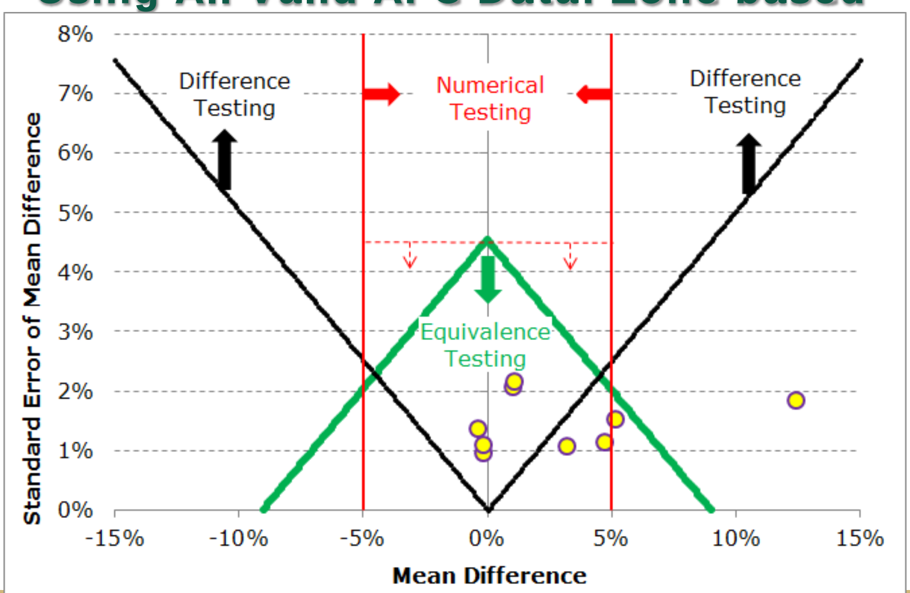
Pre-Selection of Acceptance Criterion for *Numerical Testing*

- NTD has proposed acceptance criterion $\delta = 5\%$
- The 10% precision and 95% confidence requirements are statistical in nature but this test is non-statistical
- Passing test does not ensure meeting the requirements
- Highly encourage using all valid APC data
  - Consider cap standard error of mean difference at 4.5%
- If using random sampling for estimation:
  - Minimum sampling precision for sampling plans MUST be increased to 10% minus actual $|\bar{d}|$ in percent terms
    - If actual $|\bar{d}|$ =3%, sampling precision must be increased to 7%. As a result, annual sample size would double over what 10% sampling precision would require.
  - Consider cap standard error of mean difference at 3%

Application to Certifying APC Data

19

Using All Valid APC Data: Zone based



Application to Certifying APC Data

20

Using All Valid APC Data:
Confidence Interval Based

Difference Testing					
Agency	Mean Difference	Standard Error	Lower Bound	Upper Bound	H_0 Not Rejected?
1	3.23%	1.05%	1.1%	5.3%	No
2	5.16%	1.52%	2.1%	8.2%	No
3	-0.12%	0.95%	-2.0%	1.8%	Yes
4	1.06%	2.06%	-3.0%	5.1%	Yes
5	4.74%	1.13%	2.5%	7.0%	No
6	12.42%	1.83%	8.8%	16.1%	No
7	1.10%	2.14%	-3.1%	5.3%	Yes
8	-0.36%	1.35%	-3.0%	2.3%	Yes
9	-0.15%	1.08%	-2.3%	2.0%	Yes

Using All Valid APC Data:
Confidence Interval Based (θ = 9%)

Equivalence Testing					
Agency	Mean Difference	Standard Error	Lower Bound	Upper Bound	Equivalent?
1	3.23%	1.05%	1.5%	5.0%	Yes
2	5.16%	1.52%	2.6%	7.7%	Yes
3	-0.12%	0.95%	-1.7%	1.5%	Yes
4	1.06%	2.06%	-2.4%	4.5%	Yes
5	4.74%	1.13%	2.9%	6.6%	Yes
6	12.42%	1.83%	9.4%	15.5%	No
7	1.10%	2.14%	-2.5%	4.7%	Yes
8	-0.36%	1.35%	-2.6%	1.9%	Yes
9	-0.15%	1.08%	-1.9%	1.6%	Yes

Using Random Sampling:
Acceptance criterion θ = 6%

Equivalence Testing

Agency	Mean Difference	Standard Error	Lower Bound	Upper Bound	Equivalent?	Sampling Precision
1	3.23%	1.05%	1.5%	5.0%	Yes	6.77%
2	5.16%	1.52%	2.6%	7.7%	**No**	N/A
3	-0.12%	0.95%	-1.7%	1.5%	Yes	9.88%
4	1.06%	2.06%	-2.4%	4.5%	Yes	8.94%
5	4.74%	1.13%	2.9%	6.6%	**No**	N/A
6	12.42%	1.83%	9.4%	15.5%	No	N/A
7	1.10%	2.14%	-2.5%	4.7%	Yes	8.90%
8	-0.36%	1.35%	-2.6%	1.9%	Yes	9.64%
9	-0.15%	1.08%	-1.9%	1.6%	Yes	9.85%

Application to Certifying APC Data

23

Using Random Sampling:
Acceptance criterion δ = 5%

Numerical Testing

Agency	Mean Difference	Equivalent?	Sampling Precision
1	3.23%	Yes	6.77%
2	5.16%	No	N/A
3	-0.12%	Yes	9.88%
4	1.06%	Yes	8.94%
5	4.74%	Yes	5.26%
6	12.42%	No	N/A
7	1.10%	Yes	8.90%
8	-0.36%	Yes	9.64%
9	-0.15%	Yes	9.85%

Application to Certifying APC Data

24

- Newly proposed APC rules do not ensure 10% precision & 95% confidence levels
- To protect the integrity of these statistical requirements:
  - Require using *Equivalence Testing*
  - Encourage using all valid APC data over random sampling for estimation
  - When using all valid APC data, acceptance criterion θ = 9%
  - When using random sampling:
    - Use a stricter acceptance criterion θ = 6%
    - Require agencies to account for measurement error of APC data in sampling plans
    - sampling precision = 10% - absolute value of actual mean difference

Comments on FR Notice

25

- Newly proposed *Numerical Testing* does not ensure 10% precision & 95% confidence levels
- If stay with *Numerical Testing*, improve it with the following additional requirements:
  - Encourage using all valid APC data over random sampling for estimation
  - When using all valid APC data, cap mean difference at = 4.5%
  - When using random sampling:
    - Cap mean difference at = 3%
    - Require agencies to account for measurement error of APC data in sampling plans
    - sampling precision = 10% - absolute value of actual mean difference

Comments on FR Notice

26

Having an Excel template for agency to use

- Agencies enter paired data and method for estimation
- Results from the template:
  - Certification outcome
  - If failing, indicate reasons for the failure:
    - Mean difference too big
    - Differences vary too much (i.e., standard error of mean difference too big)
    - Both are too big
  - Produce minimum sampling precision if using random sampling

Comments on FR Notice

27

For additional questions or comments, contact me:

– E-mail: xchu@cutr.usf.edu
– Phone: 813-974-9831

Contact Info

28