Understanding Small-Sample Statistics for Experimental Analysis

Sean P. Sullivan
Department of Economics & School of Law
University of Virginia

March 28, 2011

Overview

Deficiencies in Current Research

1. Generally understood that small samples are a problem
2. Not generally understood why they are a problem
3. Coordinate on certain favorite techniques
4. Favorite techniques are not always appropriate

Overview

Presentation Overview

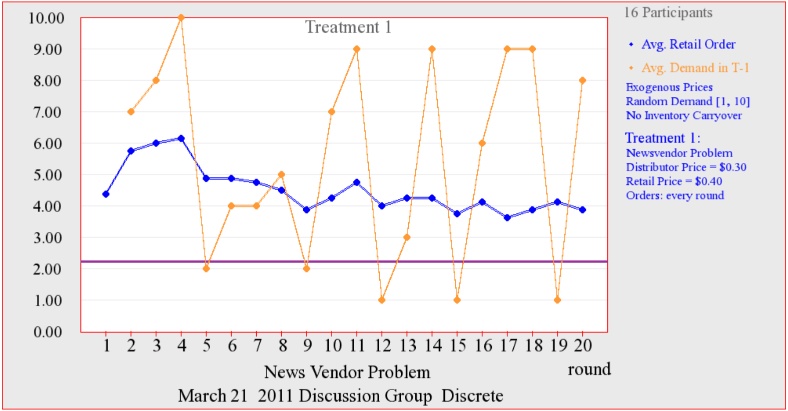
1. Intuition behind problem and solutions
2. Compare alternative tests
3. Application to Newsvendor Problem

Outline

1 Newsvender Problem
2 Normal Theory
3 Fisher Permutation
4 Wilcoxon Signed-Rank
5 Sign Test
6 Comments

Newsvender Problem

Figure: Newsvender Problem Graph



Newsvender Problem Data

Average Quantity Supplied by Participant

$$q_1 = 4.00 \qquad q_5 = 3.85$$
$$q_2 = 4.20 \qquad q_6 = 4.65$$
$$q_3 = 5.75 \qquad q_7 = 3.60$$
$$q_4 = 5.60 \qquad q_8 = 4.35$$

Data Properties

- independent ✓
- identically distributed ✓
- small sample, $n = 8$ ✓

Research Question

Summary of Current Data

$$\overline{q} = \frac{1}{n}\sum_{i=1}^{n} q_i = 4.5 \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(q_i - \overline{q})^2 = 0.792$$

Inferential Question

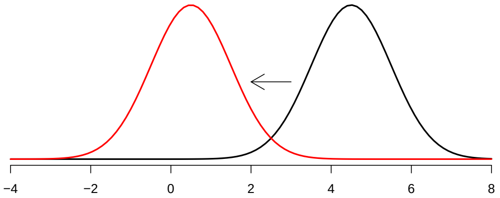
If another group of subjects played the same game tomorrow, do these data suggest tomorrow's average quantity supplied would exceed 4 units?

Simplifying Trick

Transformation

Let $x_i = q_i - 4$ as follows:

$$
\begin{array}{llcll}
q_1 = 4.00 & q_5 = 3.85 & \rightarrow & x_1 = 0.00 & x_5 = -0.15 \\
q_2 = 4.20 & q_6 = 4.65 & \rightarrow & x_2 = 0.20 & x_6 = 0.65 \\
q_3 = 5.75 & q_7 = 3.60 & \rightarrow & x_3 = 1.75 & x_7 = -0.40 \\
q_4 = 5.60 & q_8 = 4.35 & \rightarrow & x_4 = 1.60 & x_8 = 0.35
\end{array}
$$



Outline

1. Newsvender Problem
2. Normal Theory
3. Fisher Permutation
4. Wilcoxon Signed-Rank
5. Sign Test
6. Comments

Test Statistic

Familiar Normal-Theory Test Statistic

$$T_{STT} = \frac{\overline{x}}{\sqrt{V[\overline{x}]}} = \frac{\overline{x}}{s/\sqrt{n}}$$

$$\begin{aligned} \overline{x} &= 0.5 \\ s &= 0.792 \\ n &= 8 \\ T^*_{STT} &= 1.786 \end{aligned}$$

Null Distribution

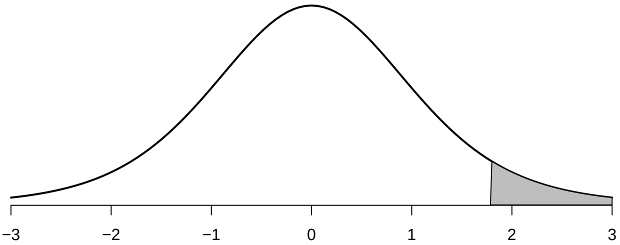
Normal Distribution, Any Sample Size
: If we assume the sample observations follow a normal distribution so that $x_1, \ldots, x_n \overset{\text{iid}}{\sim} N(0, \sigma^2)$ then $T_{STT} \sim t_{n-1}$.

Any Distribution, Large Sample Size
: As the sample size becomes large so that $n \to \infty$, for any mean-zero distribution of sample observations, $T_{STT} \overset{\text{d}}{\to} t_{n-1} \overset{\text{d}}{\to} N(0, 1)$.

Null Distribution

Calculating the P-Value

$$\begin{aligned} \text{p-value} &= P(T_{STT} \geq T^*_{STT}) \\ &= P(t_7 \geq 1.786) \\ &= 0.0586 \end{aligned}$$



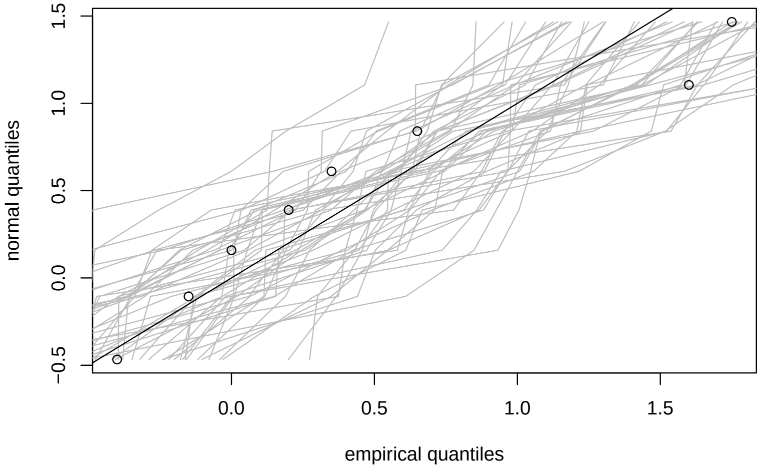
Null Distribution

Non-Normal Distribution, Small Sample Size

1. When skewness or kurtosis are severe, results of the t test can be misleading.
2. The sample size n needed to achieve good results increases as the sample distribution becomes less normal.
3. The t test is rather robust for validity, but not generally robust for efficiency.

Checking Normality Assumption

Figure: Normal Probability (Probit) Plot



Options when Normality Violated

1. More robust test statistics
2. Transformations of sample data
3. Nonparametric test statistics ✓

Outline

1. Newsvender Problem
2. Normal Theory
3. Fisher Permutation
4. Wilcoxon Signed-Rank
5. Sign Test
6. Comments

Test Statistic

Fisher Permutation Test Statistic

$$T_{FPT} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x} \quad \propto \quad T_{STT}$$

$$T^*_{FPT} = \overline{x} = 0.5$$

Deriving the Null Distribution

Symmetric Distribution
If the sampling distribution of the observations is symmetric around zero, then any observation in the sample was equally likely to have been drawn with either a positive or negative sign.

Sign Assignments
The particular signs observed in the sample are only one out of $2^n = 256$ equally likely ways that signs could have attached to the observed data.

Test Statistic Values
Each vector of signs yields a test statistic with a 1/256 chance of being observed. The full set of 256 test statistic values is the null distribution.

Deriving the Null Distribution

Example Construction of Null Distribution for $n = 3$

Consider only the last three observations: 0.65, -0.40, 0.35.

x_6	x_7	x_8	$\rightarrow$	$\bar{x}$
0.65	0.4	0.35	$\rightarrow$	0.466
0.65	0.4	-0.35	$\rightarrow$	0.233
0.65	-0.4	0.35	$\rightarrow$	0.200
0.65	-0.4	-0.35	$\rightarrow$	-0.033
-0.65	0.4	0.35	$\rightarrow$	0.033
-0.65	0.4	-0.35	$\rightarrow$	-0.200
-0.65	-0.4	0.35	$\rightarrow$	-0.233
-0.65	-0.4	-0.35	$\rightarrow$	-0.466

Null Distribution

Calculating the P-Value

$$\begin{aligned}\text{p-value} &= P(T_{FPT} \geq T^*_{FPT}) \\ &= \#\{T_{FPT} \geq T^*_{FPT}\}/2^n \\ &= 18/256 \\ &= 0.0703125\end{aligned}$$

Comments

1. No closed-form, recursive, or approximate definition of null distribution
2. Computation cost exponential in sample size
3. Bootstrapping null distribution can reduce computation cost for large samples
4. Asymptotically equivalent to the usual t test

Outline

1. Newsvender Problem
2. Normal Theory
3. Fisher Permutation
4. Wilcoxon Signed-Rank
5. Sign Test
6. Comments

Test Statistic

Signed-Rank Definition

Define SR_i as the signed-rank of observation x_i in sample x:

x_i	0	0.2	1.75	1.6	-0.15	0.65	-0.4	0.35
SR_i	1	3	8	7	-2	6	-5	4

Wilcoxon Signed-Rank Test Statistic

$$T_{WSR} = \sum_{i=1}^{n} SR_i$$

$$T^*_{WSR} = 1 + 3 + 8 + 7 - 2 + 6 - 5 + 4 = 22$$

Deriving the Null Distribution

Symmetric Distribution
: If the sampling distribution of the observations is symmetric around zero, then any observation in the sample was equally likely to have been drawn with either a positive or negative sign.

Sign Assignments
: The particular signs observed in the sample are only one out of $2^n = 256$ equally likely ways that signs could have attached to the observed data.

Test Statistic Values
: Each vector of signs yields a test statistic with a 1/256 chance of being observed. The full set of 256 test statistic values is the null distribution.

Deriving the Null Distribution

Example Construction of Null Distribution for $n = 3$

Consider only the last three observations: 0.65, -0.40, 0.35.
The signed-rank transformation of this sample is: 3,-2, 1.

SR_1	SR_2	SR_3	$\rightarrow$	$\sum SR_i$
3	2	1	$\rightarrow$	6
3	2	-1	$\rightarrow$	4
3	-2	1	$\rightarrow$	2
3	-2	-1	$\rightarrow$	0
-3	2	1	$\rightarrow$	0
-3	2	-1	$\rightarrow$	-2
-3	-2	1	$\rightarrow$	-4
-3	-2	-1	$\rightarrow$	-6

Null Distribution

Calculating the P-Value

$$\begin{aligned} \text{p-value} &= P(T_{WSR} \geq T^*_{WSR}) \\ &= \#\{T_{WSR} \geq T^*_{WSR}\}/2^n \\ &= 19/256 \\ &= 0.07421875 \end{aligned}$$

Alternative Null Distributions

Non-Data Distributions

- If no ties or zeros, null distribution can be computed without reference to data
- Pre-computed tables for small sample sizes
- Recursive algorithms for larger sample sizes

Approximate Distributions

- Same restrictions on ties and zeros
- Asymptotic approximation to normal distribution

$$\frac{T^*_{WSR}}{\sqrt{\frac{n(n+1)(2n+1)}{6}}} \xrightarrow{\mathrm{d}} N(0,1)$$

Comments

1. Popular by historical accident: alternative null distribution helpful when no computers
2. Asymptotic null distribution saves computation cost for large samples
   1. But with large samples, t test often works well anyway
3. Ties and zero values can be a headache unless null distribution is constructed by exhaustive permutation
4. Less sensitive to outliers than t test or Fisher Permutation test
5. Pretty good asymptotic relative efficiency compared to t test:
   - For normal distribution about $\frac{3}{\pi} = 95.5\%$

Outline

1 Newsvender Problem

2 Normal Theory

3 Fisher Permutation

4 Wilcoxon Signed-Rank

5 Sign Test

6 Comments

Test Statistic

Focus on Median

Let $\mu \equiv 0$ be the hypothesized median value.

Sign Test Statistic

$$T_{ST} = \sum_{i=1}^{n} 1\{x_i > 0\}$$

x_i	0	0.2	1.75	1.6	-0.15	0.65	-0.4	0.35
$1\{x_i > 0\}$	?	1	1	1	0	1	0	1

$$T^*_{ST} = 5$$

Deriving the Null Distribution

No Ties Requirement

If no observations equal zero, then any observation in the sample was equally likely to have been drawn with either a positive or negative sign.

Distribution of Signs

The sign of each observation is a random draw from a Bernoulli distribution with parameter $\theta = \frac{1}{2}$.

Null Distribution

The total number of positive signs (a sum of Bernoulli random variables) is Binomial distributed with parameters n and $\theta = \frac{1}{2}$.

Deriving the Null Distribution

Lower-Tail Test

$$\text{p-value} = \sum_{j=1}^{T_{ST}^*} \binom{n}{k} \left(\frac{1}{2}\right)^n$$

Upper-Tail Test

$$\text{p-value} = \sum_{j=1}^{n-T_{ST}^*} \binom{n}{k} \left(\frac{1}{2}\right)^n$$

Deriving the Null Distribution

The Problem with Ties to Zero

Three common ways to deal with zero-valued observations:

1. Assign ties conservatively ✓
2. Assign ties a value of $\frac{1}{2}$
3. Drop ties from sample ✓

Null Distribution

Calculating the Conservative P-value

Treat the zero observation as a negative value.

$$\text{p-value} = \sum_{j=1}^{3} \binom{8}{k} \left(\frac{1}{2}\right)^{8}$$
$$= 0.3632813$$

Calculating the Conditional P-value

Drop the zero-value, and treat as a sample with $n = 7$.

$$\text{p-value} = \sum_{j=1}^{2} \binom{7}{k} \left(\frac{1}{2}\right)^{7}$$
$$= 0.2265625$$

Comments

1. Concerns median, not (necessarily) mean
2. Without ties to zero, no distributional requirements
3. Valid even with skewed distributions
4. Generally low efficiency
5. Asymptotically relatively efficient for very heavy-tailed distributions

Outline

1. Newsvender Problem
2. Normal Theory
3. Fisher Permutation
4. Wilcoxon Signed-Rank
5. Sign Test
6. Comments

Two-Sample Analogues

Matched-Pairs Samples
: Take the vector-difference of the samples, then perform a one-sample test.

Independent Samples
: There are two-sample analogues to each of the discussed tests:

1. t test $\rightarrow$ two-sample t test
2. Fisher Permutation test $\rightarrow$ Pitman Permutation test
3. Wilcoxon Signed-Rank test $\rightarrow$ Wilcoxon-Mann-Whitney Rank-Sum test
4. Sign test $\rightarrow$ Median test

Overview

Manski's Law of Decreasing Marginal Credibility

Tests based on stronger assumptions tend to yield sharper but less credible results.

Non-Parametric Tests

1. These are not "no assumption" or "distribution free" tests
2. Relative efficiency helps to decide between competing tests
3. No statistical test is universally appropriate