Paired t-Tests & t-CI's

Engineering Statistics II
Section 9.3

Josh Engwer

TTU

2018

PART I:

Basic Experimental Design Terminology

Experimental Design

In most aspects of life, new products/services/techniques are developed as well as current ones being regularly refined and enhanced.

To determine if a new product, service or technique is more effective than one that is considered de facto standard or “status quo”, an **experiment** must be performed in order to ascertain, based on performance measurements and the relevent subsequent statistical analyses, whether the new thing is demonstrably better than the current one.

Of course, the experiment must be carefully designed to ensure that any measured advantage of the new thing is not due to chance, bias or unexplained factors.

Basic Experimental Design Terminology

Definition

The collection of 2 samples to determine cause & effect is an **experiment**.
Each data point of a sample is called an **observation** or **measurement**.
The dependent variable to be measured is called the **response**.
The person/object upon which the response is measured is a **subject**.
The manner of sample collection & grouping is called **experimental design**.
The main characteristic distinguishing the two samples is called the **factor**.
The factor's two particular values or settings are called its two **levels**.
Each sample corresponding to a level is called a **cell** or **treatment**.
Nuisance factors are uninteresting factors that influence the response.

FACTOR:	CELL SIZE:	CELLS/TREATMENTS:
Level 1	n	$x: \quad x_1, x_2, \cdots, x_n$
Level 2	n	$y: \quad y_1, y_2, \cdots, y_n$

IMPORTANT: **Observational studies** are not experiments as they don't allow cause-and-effect conclusions to be drawn.

Basic Experimental Design Terminology

Definition

The collection of 2 samples to determine cause & effect is an **experiment**.
Each data point of a sample is called an **observation** or **measurement**.
The dependent variable to be measured is called the **response**.
The person/object upon which the response is measured is a **subject**.
The manner of sample collection & grouping is called **experimental design**.
The main characteristic distinguishing the two samples is called the **factor**.
The factor's two particular values or settings are called its two **levels**.
Each sample corresponding to a level is called a **cell** or **treatment**.
Nuisance factors are uninteresting factors that influence the response.

FACTOR:	CELL SIZE:	CELL/TREAT. MEAN:	CELL/TREAT. STD DEV:
Level 1 (x)	n	$\bar{x}$	s_1
Level 2 (y)	n	$\bar{y}$	s_2

IMPORTANT: **Observational studies** are not experiments as they don't allow cause-and-effect conclusions to be drawn.

Basic Experimental Design Terminology (Example)

FACTOR: (BULB BRAND)	CELL SIZE:	CELLS/TREATMENTS: (BULB LIFETIMES in yrs)
Level 1 (x)	5	$9.22,\ 9.07,\ 8.95,\ 8.98,\ 9.54$
Level 2 (y)	5	$8.92,\ 8.88,\ 9.10,\ 8.71,\ 8.85$

or expressed in terms of means and standard deviations:

FACTOR A: (BULB BRAND)	CELL SIZE:	CELL/TREAT. MEAN (in yrs):	CELL/TREAT. STD DEV:
Level 1 (x)	5	$\bar{x} = 9.152$	$s_1 \approx 0.2410$
Level 2 (y)	5	$\bar{y} = 8.892$	$s_2 \approx 0.1406$

$$\begin{aligned} H_0 &: \quad \mu_1 = \mu_2 \\ H_A &: \quad \mu_1 \neq \mu_2 \end{aligned} \qquad \text{where } \mu_i \equiv \left(\begin{array}{c} \text{Population Mean of all} \\ \text{Brand } i \text{ light bulbs} \end{array} \right)$$

PART II:

Paired t-Tests & Paired t-CI's

Paired t-Test for μ_D (Unknown σ_1, σ_2)

Proposition

Population:	*Two Normal Populations with unknown* σ_1, σ_2
Realized Samples:	$\mathbf{x} := (x_1, x_2, \cdots, x_n)$ *with mean* $\bar{x}$, *std dev* $s_1, n \geq 2$ $\mathbf{y} := (y_1, y_2, \cdots, y_n)$ *with mean* $\bar{y}$, *std dev* $s_2, n \geq 2$
Difference Population:	$D := X - Y \implies D \sim \text{Normal}(\mu_D, \sigma_D^2)$ *where* $\mu_D = \mu_1 - \mu_2$ & $\sigma_D^2 = \sigma_1^2 + \sigma_2^2 - 2 \cdot \mathbb{C}[X, Y]$
Realized Pairs:	$(x_1, y_1), (x_2, y_2), \cdots (x_n, y_n)$
Realized Differences:	$\mathbf{d} := (d_1, d_2, \cdots, d_n)$ *where* $d_k := x_k - y_k$
Test Statistic Value: $W(\mathbf{d}; \delta_0)$	$t = \dfrac{\bar{d} - \delta_0}{s_d / \sqrt{n}}$

HYPOTHESIS TEST:	***REJECTION REGION AT LVL*** α :
$H_0 : \mu_1 - \mu_2 = \delta_0$ *vs.* $H_A : \mu_1 - \mu_2 > \delta_0$	$t \geq t^*_{n-1;\alpha}$
$H_0 : \mu_1 - \mu_2 = \delta_0$ *vs.* $H_A : \mu_1 - \mu_2 < \delta_0$	$t \leq t^*_{n-1;1-\alpha}$
$H_0 : \mu_1 - \mu_2 = \delta_0$ *vs.* $H_A : \mu_1 - \mu_2 \neq \delta_0$	$t \leq t^*_{n-1;1-\alpha/2}$ *or* $t \geq t^*_{n-1;\alpha/2}$

Paired t-Test for μ_D (Unknown σ_1, σ_2)

Proposition

Population:	*Two Normal Populations with unknown* σ_1, σ_2
Realized Samples:	$\mathbf{x} := (x_1, x_2, \cdots, x_n)$ *with mean* $\bar{x}$, *std dev* $s_1, n \geq 2$ $\mathbf{y} := (y_1, y_2, \cdots, y_n)$ *with mean* $\bar{y}$, *std dev* $s_2, n \geq 2$
Difference Population:	$D := X - Y \implies D \sim$ Normal(μ_D, σ_D^2) *where* $\mu_D = \mu_1 - \mu_2$ & $\sigma_D^2 = \sigma_1^2 + \sigma_2^2 - 2 \cdot \mathbb{C}[X, Y]$
Realized Pairs:	$(x_1, y_1), (x_2, y_2), \cdots (x_n, y_n)$
Realized Differences:	$\mathbf{d} := (d_1, d_2, \cdots, d_n)$ *where* $d_k := x_k - y_k$
Test Statistic Value: $W(\mathbf{d}; \delta_0)$	$t = \dfrac{\bar{d} - \delta_0}{s_d / \sqrt{n}}$

HYPOTHESIS TEST:	***P-VALUE DETERMINATION:***
$H_0 : \mu_D = \delta_0$ vs. $H_A : \mu_D > \delta_0$	*P-value* $= 1 - \Phi_t(t; \nu = n - 1)$
$H_0 : \mu_D = \delta_0$ vs. $H_A : \mu_D < \delta_0$	*P-value* $= \Phi_t(t; \nu = n - 1)$
$H_0 : \mu_D = \delta_0$ vs. $H_A : \mu_D \neq \delta_0$	*P-value* $= 2 \cdot [1 - \Phi_t(\lvert t \rvert; \nu = n - 1)]$

Paired t-CI for μ_D (Unknown σ_1, σ_2)

Proposition

Given two normal populations with means μ_1 and μ_2.
Let $x_1, x_2, \cdots, x_n$ be a sample taken from the 1^{st} population.
Let $y_1, y_2, \cdots, y_n$ be a sample taken from the 2^{nd} population.
Moreover, suppose the two samples are paired as follows:

$$(x_1, y_1),\ (x_2, y_2),\ \cdots,\ (x_n, y_n)$$

where the pairwise differences $(d_k := x_k - y_k)$ are: $\quad d:\ d_1, d_2, \cdots, d_n$

*Then the $100(1-\alpha)\%$ **paired t-CI for** $\mu_D := \mu_1 - \mu_2$ is*

$$\left(\overline{d} - t^*_{n-1;\alpha/2} \cdot \frac{s_d}{\sqrt{n}},\ \overline{d} + t^*_{n-1;\alpha/2} \cdot \frac{s_d}{\sqrt{n}}\right)$$

—- OR WRITTEN MORE COMPACTLY —-

$$\overline{d} \pm t^*_{n-1;\alpha/2} \cdot \frac{s_d}{\sqrt{n}}$$

Computing s_d^2 when given Means $\bar{x}, \bar{y}$, Std Dev's s_x, s_y and either Covariance s_{xy} or Correlation r_{xy}

When the common sample size n is very large (i.e. $n \gg 20$), it's too tedious to compute means and std deviations from the samples themselves.

Instead, a research paper, for example, would provide the following:

- Sample means $\bar{x}, \bar{y}$
- Standard Deviations s_x, s_y
- Either Covariance s_{xy} or Correlation r_{xy}

Here's how to compute the realized differences' mean & std dev then:

- Mean $$\bar{d} := \underbrace{\frac{1}{n}\sum_{k=1}^{n} d_k}_{\text{Given samples}} = \underbrace{\bar{x} - \bar{y}}_{\text{Given means}}$$
- Variance $$s_d^2 := \underbrace{\frac{1}{n-1}\sum_{k=1}^{n}\left(d_k - \bar{d}\right)^2}_{\text{Given samples}} = \underbrace{s_x^2 + s_y^2 - 2s_{xy}}_{\text{Given covariance}} = \underbrace{s_x^2 + s_y^2 - 2r_{xy}s_x s_y}_{\text{Given correlation}}$$

PART III:

When to Choose Dependent (Paired) Samples vs. Independent Samples

Dependent Samples Versus Independent Samples

The variance of paired difference rv helps inform the better choice:

$$\mathbb{V}\left[\overline{D}\right] := \mathbb{V}\left[\overline{X} - \overline{Y}\right] = \mathbb{V}\left[\overline{X}\right] + \mathbb{V}\left[\overline{Y}\right] - 2 \cdot \mathbb{C}\left[\overline{X}, \overline{Y}\right] = \frac{2\sigma^2(1-\rho)}{n}$$

- If the entire body of subjects in the experiment is largely heterogeneous and the pairing scheme chosen is reasonable such that any extraneous factors are similar between the two members in each pair (which suggests each pair's members bear small variance but large correlation), then use the paired t-test as $\mathbb{V}\left[\overline{D}\right]$ will be smaller $\implies$ smaller ESE.
  - A blatant dependent sample would be observing/measuring a subject pre-treatment and then post-treatment.
  - Naturally correlated pairs include twins, father and son, two cats from the same litter, two identical genetically cloned plants, etc...
- If the entire body of subjects in the experiment is largely homogeneous, then pairing will not be very effective – the tiny gains from pairing will be negated by the loss of half the degrees of freedom. Therefore, use two independent sample t-test instead.

Choosing independent or dependent samples appropriately given the particular situation will result in narrower CI's and more-powerful tests.

Dependent Samples Versus Independent Samples

Both independent t-CI's and paired t-CI's have the same general form:

$$(\bar{x} - \bar{y}) \pm t^*_{\nu;\alpha/2} \cdot \text{ESE}$$

What differs are the est. std. error (ESE) & degrees of freedom (ν):

	INDEPENDENT SAMPLES: (equal sample size, n)	**DEPENDENT SAMPLES:** (n pairs)
ESE $(\widehat{\sigma})$	$s_{pooled} \cdot \sqrt{\frac{2}{n}}$	$s_d \cdot \sqrt{\frac{1}{n}}$
dof's (ν)	$2(n-1)$	$n-1$

For equal ESE's, fewer degrees of freedom (dof's) results in:

- Wider CI's
- Less-powerful tests

In general, for large samples $(n > 40)$, a smaller ESE is worth having only half the degrees of freedom, hence prefer using paired t-tests unless the entire sample is extremely similar or homogeneous.

Textbook Logistics for Section 9.3

CONCEPT	TEXTBOOK NOTATION	SLIDES/OUTLINE NOTATION
Expected Value of rv	$E(X)$	$\mathbb{E}[X]$
Variance of rv	$V(X)$	$\mathbb{V}[X]$
Covariance of rv's	$\text{Cov}(X, Y)$	$\mathbb{C}[X, Y]$
Correlation of rv's	$\rho_{X,Y}$	ρ_{XY}
Sample Variance	S_{xx}, S_{yy}	s_x^2, s_y^2
Sample Covariance	S_{xy}	s_{xy}
Sample Correlation	r	r_{xy}
t-Cutoff	$t_{\alpha/2,\nu}$	$t^*_{\nu;\alpha/2}$
Null Hypothesis Value	Δ_0	δ_0
Realized Differences Std Dev	s_D	s_d

Fin.