**Hypothesis Tests for Comparisons**

**Problem** A company is considering two procedures for measuring the mass of gizmos produced on an assembly line. It would like to know whether, on average, the two procedures produce the same results. It gathers data by randomly selecting 50 gizmos off the assembly line, measuring each of their masses using both procedures. How do we determine whether this data provides strong evidence that, on average, the two procedures produce different results?

**The parameter**  = average (procedure1 – procedure2) for all gizmos

**The hypotheses** H0 :  = 0

Ha :  ≠ 0

**The data** 50 differences in measurements d = x1 – x2

**The sample statistics** = average of the 50 measurement differences

sd = standard deviation of the 50 measurement differences

**The sampling distribution** If the distribution of all possible differences is normal and the null hypothesis is true, then t = has a t-distribution with 49 degrees of freedom. If the distribution is not perfectly normal, then t = has approximately a t-distribution.

**The p-value** 2Prob(T49 ≥ |t|)

The method illustrated above is called the **paired t-test**, since the data is paired.

Suppose = 0.14 gm and sd = 0.13 gm. Then t = = 7.62,

> 1-pt(7.62,49)

[1] 3.656297e-10

**So, the p-value = 2\*0 = 0. So, there is strong evidence that the two procedures produce different results on average.**

**General Format for a paired t-test**

**Hypotheses**

H0 : = 

Ha : > or < or ≠

**Data** d1, d2, …, dn paired differences

**Test statistic** t =  has (approximately) a t-distribution with n-1 degrees

of freedom. (Where is the average of the paired differences and sd is the standard deviation of the paired differences.)

**p-value** Prob(Tn-1 ≥ t) or Prob(Tn-1 ≤ t) or 2 Prob(Tn-1 ≥| t|) depending on

the alternative hypothesis.

**Example** On average, is a student’s MathSAT score higher than the student’s VerbalSAT score?

If = average difference MathSAT – VerbalSAT

H0 : = 0

Ha : > 0

The data-frame Students contains values for 200 students

> dbar<-mean(~(MathSAT-VerbalSAT),data = Students)

> sdd<-sd(~(MathSAT-VerbalSAT),data = Students)

> dbar

[1] 20.195

> sdd

[1] 72.19437

> t<-(dbar-0)/(sdd/sqrt(200))

> t

[1] 3.955993

> 1-pt(t,199)

[1] 5.298035e-05

**With a p-value this small, the data is very strong evidence that, on average, a student’s MathSAT score is higher that the student’s VerbalSAT.**

**Letting R to do all of the work**

Suppose we want to do a paired t-test where the variables are in the column COL1 and COL2 of the data-frame FRAME. The difference of interest is COL1 – COL2.

t.test(FRAME$COL1 , FRAME$COL2 , alternative = “ “ , paired = TRUE)

**Apply to the example above**:

> t.test(Students$MathSAT,Students$VerbalSAT, alternative = "greater", paired = TRUE)

Paired t-test

data: MathSAT and VerbalSAT

t = 3.956, df = 199, p-value = 5.298e-05

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:

11.75889 Inf

sample estimates:

mean of the differences

20.195

**A New Problem**

Suppose the company has two different assembly lines that produce gizmos. It would like to know whether the average mass of the gizmos produced by the line 1 is the same as the average mass of the gizmos produced by line 2.

**The Parameters**

 = average mass for assembly line 1

 = average mass for assembly line 2

**The Hypotheses**

H0 :  = 

Ha :  ≠ 

**The Data**

Two **independent** samples of gizmos are selected, a sample of size n1 from line1 and a sample of size n2 from line2. The masses for line1 samples are and the masses for line2 are.

**The Sample Statistics**

, , , 

**The sampling distribution**

t =  has (approximately) a t-distribution with



**Rule of thumb df = smaller of n1-1 and n2 -1**.

**P-value**

Prob(T\* ≥ t) or Prob(T\* ≤ t) or 2 Prob(T\* ≥|t|) depending on

the alternative hypothesis.

**This procedure is called the two (independent) sample t-test. It is in contrast to the paired t-test.**

**Example**

**H0 :** = 

Ha : > 

A sample of size 20 produces = 4.0 and s1 = 0.5. An independent sample of size 30 produces = 3.9 and s2 = 0.4. Find the value of the test statistic and the p-value.

t = = 0.75

Using the rule of thumb, df = 19. From R

> 1-pt(.75,19)

[1] 0.2312208

So, with a p-value of 23%, we keep the null hypothesis.

**Using R**

If the two samples are contained in the columns COL1 of the data-frame FRAME1 and COL2 of the data-frame FRAME2

**t.test(FRAME1$COL1, FRAME2$COL2, , alternative = ” “** )

**Example.** The data-frame Salaries contains annual salaries (in $1,000) for 50 individuals with gender1 and 50 individuals with gender2. The data are contained in the columns Salary1 and Salary2. Use this data to test whether there is a difference between the average annual salary for all gender1’s and the average annual salary for all

gender 2’s. **We assume that the 50 gender1’s behind the data are independent of the 50 gender2’s behind the data. If the 50 gender1’s were the spouses of the 50 gender2’s the sample would not be independent and the two-sample procedure would not be appropriate.**

**The parameters.**

1 = average annuak salary of all individuals with gender1

2 = average annual salary of all individuals with gender2

The Hypotheses

H0 : 1 = 2

Ha :  ≠ 

**Letting R do all the work**

> t.test(Salaries$Salary1,Salaries$Salary2 ,alternative = "two.sided")

Welch Two Sample t-test

data: Salary1 and Salary2

t = -2.658, df = 77.064, **p-value = 0.009555**

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-38.108788 -5.465212

sample estimates:

mean of x mean of y

41.631 63.418

**With such a small p-value, the evidence is very strong that there is a difference between the average for gender1 and the average for gender2.**

**Example** The data-frame **Students** has a column **hsGPA** that contains student high school GPA’s for 1000 college students. The data-frame **StudentSurvey** has a column **GPA** that contains college GPA’s for 362 college students. Apply the appropriate t-test to determine whether the average college GPA for all college students is less than the average high school GPA for all college students.

**Is a two independent sample t-test or a paired t-test appropriate?**

**Since the samples are independent of each other, we apply the two-sample t-test.**

**Define the parameters and state the hypotheses**

**μ1. = average college GPA score for all college students**

**μ2 = average high school for all college students**

**H0 : μ1 = μ2. Ha : μ1 < μ2**

**Use R to find the p-value and answer the question**

> t.test(StudentSurvey$GPA,Students$hsGPA, alternative = "less")

Welch Two Sample t-test

data: StudentSurvey$GPA and Students$hsGPA

t = -14.725, df = 717.25, p-value < 2.2e-16

alternative hypothesis: true difference in means is less than 0

95 percent confidence interval:

-Inf -0.344359

sample estimates:

mean of x mean of y

3.157942 3.545669

**The p-value is essentially 0. So, we conclude that the average college GPA for all college students is less than the average high school GPA for all college students.**

**Exercises 19**

1. Load the packages: MASS and Lock5withR. The data-frame **survey (**in the MASS package) contains data from 237 students at the University of Adelaide in Australia. One of the variables in this data-frame is **Pulse** (pulse rate in beats per minute.). The data frame **StudentSurvey** (in the Lock5withR package) contains data for 362 American college students. It also has a variable **Pulse**. Is there a difference between the average pulse rate of Australian students and the average pulse rate of American students?
2. Is a paired t.-test or a two independent sample t-test the appropriate procedure to use?
3. Use the data in these two data frames to determine whether there is strong evidence for s difference. Include the definitions of the parameters, the hypotheses, and the R commands you used in your solution.
4. The data-frame **StatGrades** (in the Lock5withR package) contains data on exam/test performance of 50 students in an intro statistics course. Two of the variables are **Exam1** (score on 1st test) and **Exam2** (score on 2nd test). Does this data provide strong evidence that, on average, a student in an intro stats course does better on the second test that on the first test?
5. Is a paired t-test or a two independent sample t-test the appropriate procedure to use?
6. Apply the appropriate procedure to get the p-value and answer the question. Include the definitions of the parameters, the hypotheses, and the R commands you used in your solution.
7. **(6 pts)** The data-frame **SnowGR** contains, for each month of the years (1893 – 2012), the amount of snow (inches) that Grand Rapids received in that month. Assume this data is a random sample of snowfall figures for all years from 1 AD to 3000 AD. In an average year does GR receive more snowfall in January (variable = **Jan**) than in February (variable = **Feb**)?
8. Is a paired t-test or a two independent sample t-test the appropriate procedure to use?
9. Apply the appropriate procedure to get the p-value and answer the question. Include the definitions of the parameters, the hypotheses, and the R commands you used in your solution.