Paired t-test

STAT 401 - Statistical Methods for Research Workers

Jarad Niemi

Iowa State University

6 September 2013

Cedar-apple rust

Cedar-apple rust is a (non-fatal) disease that affects apple trees. Its most obvious symptom is rust-colored spots on apple leaves. Red cedar trees are the immediate source of the fungus that infects the apple trees. If you could remove all red cedar trees within a few miles of the orchard, you should eliminate the problem. In the first year of this experiment the number of affected leaves on 8 trees was counted; the following winter all red cedar trees within 100 yards of the orchard were removed and the following year the same trees were examined for affected leaves.

Cedar-apple rust

Cedar-apple rust is a (non-fatal) disease that affects apple trees. Its most obvious symptom is rust-colored spots on apple leaves. Red cedar trees are the immediate source of the fungus that infects the apple trees. If you could remove all red cedar trees within a few miles of the orchard, you should eliminate the problem. In the first year of this experiment the number of affected leaves on 8 trees was counted; the following winter all red cedar trees within 100 yards of the orchard were removed and the following year the same trees were examined for affected leaves.

- Statistical hypothesis:
  - H_0: Removing red cedar trees increases or maintains the same mean number of rusty leaves.
  - H_1: Removing red cedar trees decreases the mean number of rusty leaves.

Cedar-apple rust

Cedar-apple rust is a (non-fatal) disease that affects apple trees. Its most obvious symptom is rust-colored spots on apple leaves. Red cedar trees are the immediate source of the fungus that infects the apple trees. If you could remove all red cedar trees within a few miles of the orchard, you should eliminate the problem. In the first year of this experiment the number of affected leaves on 8 trees was counted; the following winter all red cedar trees within 100 yards of the orchard were removed and the following year the same trees were examined for affected leaves.

- Statistical hypothesis:
  - H_0: Removing red cedar trees increases or maintains the same mean number of rusty leaves.
  - H_1: Removing red cedar trees decreases the mean number of rusty leaves.

- Statistical question:

  What is the reduction of rusty leaves in our sample between year 1 and year 2 (perhaps due to removal of red cedar trees?

Data

Here are the data

	year1	year2	diff
1	38	32	6
2	10	16	-6
3	84	57	27
4	36	28	8
5	50	55	-5
6	35	12	23
7	73	61	12
8	48	29	19

Assumptions

Let

- Y_{1j} be the number of rusty leaves on tree j in year 1
- Y_{2j} be the number of rusty leaves on tree j in year 2

Assumptions

Let

- Y_{1j} be the number of rusty leaves on tree j in year 1
- Y_{2j} be the number of rusty leaves on tree j in year 2

Assume

$$D_j = Y_{1j} - Y_{2j} \overset{iid}{\sim} N(\mu, \sigma^2)$$

Assumptions

Let

- Y_{1j} be the number of rusty leaves on tree j in year 1
- Y_{2j} be the number of rusty leaves on tree j in year 2

Assume

$$D_j = Y_{1j} - Y_{2j} \overset{iid}{\sim} N(\mu, \sigma^2)$$

Then

H_0: $\mu = 0$ $(\mu \leq 0)$

H_1: $\mu > 0$

Pvalue

Test statistic

$$t = \frac{\overline{D_j} - \mu}{SE(\overline{D_j})}$$

Pvalue

Test statistic

$$t = \frac{\overline{D_j} - \mu}{SE(\overline{D_j})}$$

where $SE(\overline{D_j}) = s/\sqrt{n}$ with

- n being the number of observations (differences) and
- s being the sample standard deviation of the differences.

Pvalue

Test statistic

$$t = \frac{\overline{D_j} - \mu}{SE(\overline{D_j})}$$

where $SE(\overline{D_j}) = s/\sqrt{n}$ with

- n being the number of observations (differences) and
- s being the sample standard deviation of the differences.

If H_0 is true, then $\mu = 0$ and $t \sim t_{n-1}$.

Pvalue

Test statistic

$$t = \frac{\overline{D_j} - \mu}{SE(\overline{D_j})}$$

where $SE(\overline{D_j}) = s/\sqrt{n}$ with

- n being the number of observations (differences) and
- s being the sample standard deviation of the differences.

If H_0 is true, then $\mu = 0$ and $t \sim t_{n-1}$. The pvalue is $P(t_{n-1} > t)$ since this is a one-sided test.

Pvalue

Test statistic

$$t = \frac{\overline{D_j} - \mu}{SE(\overline{D_j})}$$

where $SE(\overline{D_j}) = s/\sqrt{n}$ with

- n being the number of observations (differences) and
- s being the sample standard deviation of the differences.

If H_0 is true, then $\mu = 0$ and $t \sim t_{n-1}$. The pvalue is $P(t_{n-1} > t)$ since this is a one-sided test.

For these data,

- $\overline{D_j} = 10.5$
- $SE(\overline{D_j}) = 4.31$
- $t = 2.43$
- $p = 0.02$

Confidence interval

The 100(1-α)% confidence interval has lower endpoint

$$\overline{D_j} - t_{n-1}(1-\alpha)SE(\overline{D_j})$$

and upper endpoint at infinity

Confidence interval

The 100(1-α)% confidence interval has lower endpoint

$$\overline{D_j} - t_{n-1}(1-\alpha)SE(\overline{D_j})$$

and upper endpoint at infinity

For these data at 95% confidence, the lower endpoint is

$$10.5 - 1.89 \cdot 4.31 = 2.33$$

Confidence interval

The 100(1-α)% confidence interval has lower endpoint

$$\overline{D_j} - t_{n-1}(1-\alpha)SE(\overline{D_j})$$

and upper endpoint at infinity

For these data at 95% confidence, the lower endpoint is

$$10.5 - 1.89 \cdot 4.31 = 2.33$$

So we are 95% confident that the true difference in the number of rusty leaves is greater than 2.33.

SAS code for paired t-test

```
DATA leaves;
  INPUT tree year1 year2;
  DATALINES;
1 38 32
2 10 16
3 84 57
4 36 28
5 50 55
6 35 12
7 73 61
8 48 29
;

PROC TTEST DATA=leaves SIDES=U;
    PAIRED year1*year2;
    RUN;
```

The TTEST Procedure

Difference: year1 - year2

N	Mean	Std Dev	Std Err	Minimum	Maximum
8	10.5000	12.2007	4.3136	-6.0000	27.0000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
10.5000	2.3275	Infty	12.2007	8.0668	24.8317

DF	t Value	Pr > t
7	2.43	0.0226

Conclusion

Removal of red cedar trees within 100 yards is associated with a significant reduction in rusty apple leaves (paired t-test t=2.43, p=0.023).

Conclusion

Removal of red cedar trees within 100 yards is associated with a significant reduction in rusty apple leaves (paired t-test t=2.43, p=0.023). The mean reduction in rust color leaves is 10.5 [95% CI $(2.33, \infty)$].