# Statistics for Computing

## MA4413 Lecture 10A

Kevin O'Brien

Kevin.obrien@ul.ie

Dept. of Mathematics & Statistics,
University *of* Limerick

Autumn Semester 2013

# Today's Class

1. Paired t-test
2. Difference of two mean (large samples)
3. Difference of two mean (small samples)
4. Difference of two proportions

# Paired Data

- Two measurements are paired when they come from the same case (person, item, observational unit). It is not ncessary for the measurements to be denominated in the same units, but very helpful.
- Pairing is determined by a study's design and the way the data values are obtained, and with the actual data values themselves not being particularly relevant.
- Observations are paired rather than independent when there is a natural link between an observation in one set of measurements and a particular observation in the other set of measurements.
- Examples of paired data: ***before and after*** measurements, ***with and without*** measurements, and two simultaneous measurements on the same item.

# Paired Data

- We are usually required to compute the case-wise difference for each data pairing.
- Importantly, although we start out with two samples of data, we can look at the data as a single sample of ***case-wise differences***.

$$d_i = x_i - y_i$$

- We can use the same methodologies that we have encountered previously for making decisions based on paired data.
- (Remark: For most paired data studies, the sample sizes are very small.)

# The Paired t-test

- We will often be required to compute the case-wise differences, the average of those differences and the standard deviation of those difference.
- The mean difference for a set of differences between paired observations is

$$\bar{d} = \frac{\sum d_i}{n}$$

- The computational formula for the standard deviation of the differences between paired observations is

$$s_d = \sqrt{\frac{\sum d_i^2 - n\bar{d}^2}{n-1}}$$

# The Paired t-test

Using the sample to make inferences about the general population of case-wise differences.

- Often we are making conclusions for the population of differences. (Is a training regime effective? - based on a paired data sample.)
- Let $\mu_d$ be mean value for the population of case-wise differences.
- The null hypothesis is that that $\mu_d = 0$ (i.e. no difference)
- Given $\bar{d}$ mean value for the sample of differences, and $s_d$ standard deviation of the differences for the paired sample data, we can perform inference procedures as we have done previously.

# Example 1: Paired Difference (a)

- An automobile manufacturer collects mileage data for a sample of $n = 10$ cars in various weight categories using a standard grade of gasoline with and without a particular additive.
- Of course, the engines were tuned to the same specifications before each run, and the same drivers were used for the two gasoline conditions (with the driver in fact being unaware of which gasoline was being used on a particular run).
- Given the mileage data on the next slide, test the hypothesis that there is no difference between the mean mileage obtained with and without the additive, using the 5 percent level of significance
- (Remark in lecture: Enough evidence for haulage company to start buying this additive?)

# Example 1: Paired Difference (b)

| car | with additive | without additive | $d_i$ | $d_i^2$ |
|---|---|---|---|---|
| 1 | 36.7 | 36.2 | 0.5 | 0.25 |
| 2 | 35.8 | 35.7 | 0.1 | 0.01 |
| 3 | 31.9 | 32.3 | -0.4 | 0.16 |
| 4 | 29.3 | 29.6 | -0.3 | 0.09 |
| 5 | 28.4 | 28.1 | 0.3 | 0.09 |
| 6 | 25.7 | 25.8 | -0.1 | 0.01 |
| 7 | 24.2 | 23.9 | 0.3 | 0.09 |
| 8 | 22.6 | 22.0 | 0.6 | 0.36 |
| 9 | 21.9 | 21.5 | 0.4 | 0.16 |
| 10 | 20.3 | 20.0 | 0.3 | 0.09 |

# Example 1: Paired Difference (c)

- The average of the case wise differences is computed as

$$\bar{d} = \frac{\sum d_i}{n}$$

$$\bar{d} = \frac{0.5 + 0.1 - 0.4 + \ldots + 0.30}{10} = 0.17$$

- Also, using last column, $\sum d_i^2 = (0.25 + 0.01 + 0.16 + \ldots + 0.09) = 1.31$

# Example 1: Paired Difference (d)

**Sample standard deviation of the case-wise differences**:

$$s_d = \sqrt{\frac{\sum d_i^2 - n\bar{d}^2}{n-1}}$$

We know the following:

- The sample size $n$ which is 10.
- The average of the case-wise differences. $\bar{d} = 0.17$
- $\sum d_i^2 = 1.31$

# Example 1: Paired Difference (e)

**Sample standard deviation of the case-wise differences**:

$$s_d = \sqrt{\frac{\sum d_i^2 - n\bar{d}^2}{n-1}}$$

$$s_d = \sqrt{\frac{1.31 - 10(0.17)^2}{9}} = 0.337$$

**The standard error:**

$$S.E.(\bar{d}) = \frac{s_d}{\sqrt{n}} = \frac{0.337}{3.16} = 0.107$$

# Example 1: Paired Difference (f)

**Null and Alternative Hypotheses**:

- That is, the null hypothesis is:
  $H_0 : \mu_d = 0$ Additive makes no difference to performance
  $H_1 : \mu_d \neq 0$ Additive makes a significant difference to performance

**Test Statistic**:

- Test Statistic

$$TS = \frac{\bar{d} - \mu_d}{S.E.(\bar{d})} = \frac{0.17 - 0}{0.107} = 1.59$$

# Example 1: Paired Difference (g)

**Critical value**:

- $\alpha = 0.05, k = 2$
- small sample , so $df = n - 1 = 9$
- As with earlier examples in the course, CV is found to be **2.262** from the statistical tables.

**Decision Rule**:
Is $|TS| > CV$?
No, we fail to reject the null hypothesis. There is no enough evidence to suggest this additive is effective in improving mileages for automobiles.

# Example 1: Paired Difference (h)

**Confidence Interval**:

- Recall our sample mean $\bar{x}$, the standard error $S.E(\bar{x})$ and the quantile from the $t-$ distribution.
- We can use these values to compute a 95% confidence interval.
- The 95% confidence interval can be computed as $0.17 \pm (2.262 \times 0.17 = (-0.21, 0.55)$
- Notice that 0 is within that range of values. This supports our conclusion to reject the Null Hypothesis.

## Two Sample Procedures

- So far we have looked at single sample procedures.
- We can generalise our methodologies for comparing two samples.
- Our point estimates are typically differences in sample statistics. i.e.

$$\bar{X}_1 - \bar{X}_2$$

$$\hat{p}_1 - \hat{p}_2$$

- We can use these point estimates to make inference on differences for populations.

$$\mu_1 - \mu_2$$

$$\pi_1 - \pi_2$$

# Two Sample Procedures - Confidence Intervals

- When computing confidence intervals, all that is required is the calculation of the appropriate standard error value.
- The two-sample standard error calculations contain statistical information from both samples.
- See the formula sheet. (Practice with these calculaions will take place in next week's tutorials.)
- There are some minor issues that will arise with each type of procedure. These will be explained in relevant examples.

# Two Sample Procedures - small samples and degrees of freedom

- Let $n_1$ and $n_2$ be the sample sizes of two samples.
- When deciding whether to use the large sample approach or the small sample appproach, we will use the following rule of thumb:
  Small sample : $n_1 + n_2 \leq 30$
  Large sample : otherwise
- For small samples the appropriate degrees of freedom is $(n_1 - 1) + (n_2 - 1)$ i.e. $n_1 + n_2 - 2$

## Two Sample Procedures - small samples and degrees of freedom

- When performing hypothesis tests, we are usually interested in determining whether or not the population parameters can be considered equal for both populations.
- Another way of expressing this is that the difference in population parameters is 0.
- The following two hypotheses are directly equivalent.

$$H_o : \mu_1 = \mu_2$$

$$H_o : \mu_1 - \mu_2 = 0$$

- Equivalently, for proportions:

$$H_o : \pi_1 = \pi_2$$

$$H_o : \pi_1 - \pi_2 = 0$$

# Two Sample Procedures - Standard Errors

- **Small Samples**: When computing the standard error for difference in sample means - take care to use the appropriate standard error formula. (i.e using the pooled variance calculation.)
- **Hypothesis Tests for Propostions:** Use the aggregate proportion formula $\bar{p}$.
- See formula sheet.

# Example 2: Difference in Means (a)

- Two sets of patients are given courses of treatment under two different drugs.
- The benefits derived from each drug can be stated numerically in terms of the recovery times measured in days
- The sample size, mean and standard deviations for both groups are given below.

| Group | Sample Size | Mean | Std. Dev. |
|---|---|---|---|
| 1 | $n_1 = 40$ | $\bar{x}_1 = 3.3$ days | $s_1 = 1.524$ |
| 2 | $n_2 = 45$ | $\bar{x}_2 = 4.3$ days | $s_2 = 1.951$ |

# Example 2: Difference in Means (b)

- The first step in hypothesis testing is to specify the null hypothesis and an alternative hypothesis.
- When testing differences between mean recovery times, the null hypothesis is that the two population means are equal.
- That is, the null hypothesis is:
  $H_0 : \mu_1 = \mu_2$ (The population means are equal - no difference in drug treatments)
  $H_1 : \mu_1 \neq \mu_2$ (The population means are different - difference in drug treatments)

(Remark: Two-tailed Test, therefore $k = 2$, and $\alpha = 0.05$)

# Example 2: Difference in Means (c)

- **Point Estimate**: The observed difference in means is $\bar{x}_1 - \bar{x}_2 \ = \ 4.3 - 3.3 = 1$ day.
- The relevant formula for the standard error is

$$S.E(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$S.E(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{(1.524)^2}{40} + \frac{(1.951)^2}{45}}$$

$$S.E(\bar{x}_1 - \bar{x}_2) = 0.377 \text{ days}$$

# Example 2: Difference in Means (d)

- The Test statistic is therefore

$$TS = \frac{\text{Point Estimate} - \text{Difference under } H_0}{\text{Std. Error}}$$

$$TS = \frac{1-0}{0.377} = 2.65$$

# Example 2: Difference in Means (e)

- As the sample was large, we could use $CV = 1.96$ (as always, two tailed procedure, with $\alpha = 0.05$).
- **Decision Rule**: Is the $TS > CV$ ?
  Is $2.65 > 1.96$ ? - Yes , we reject the null hypothesis.
- There is enough evidence to suggest that there is a difference in drug treatments.

# Example 3: Difference in Means (a)

- For a random sample of 10 light bulbs, the mean bulb life is 4,000 hr with a standard deviation of 200 hours.
- For another brand of bulbs whose useful life is also assumed to be normally distributed, a random sample of 8 has a sample mean of 4,300 hours and a sample standard deviation of 250 hours.
- Test the hypothesis that there is no difference between the mean operating life of the two brands of bulbs, using the 5 percent level of significance.

# Example 3: Difference in Means (b)

**Summary of Information**

- $n_1 = 10$ and $n_2 = 8$.
- $\bar{x}_1 = 4000$ hours, $\bar{x}_2 = 4,300$ hours , therefore $\bar{x}_2 - \bar{x}_1 = 300$ hours
- $s_1 = 200$ hours, $s_2 = 250$ hours.
- Small aggregate sample - Degrees of freedom $n_1 + n_2 - 2 = 10 + 8 - 2 = 16$

# Example 3: Difference in Means (c)

- Look at the standard error formula for difference in means for a small aggregate sample.
- Before we compute the standard error, we must compute the **pooled variance estimate**

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$s_p^2 = \frac{(9 \times 200^2) + (7 \times 250^2)}{16}$$

$$s_p^2 = 49843.75$$

# Example 3: Difference in Means (d)

**Computing the Standard Error**

$$S.E(x_1 - x_2) = \sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$S.E(x_1 - x_2) = \sqrt{49843.75\left(\frac{1}{10} + \frac{1}{8}\right)}$$

$$S.E(x_1 - x_2) = \sqrt{11214.84} = 105.9$$

# Example 3: Difference in Means (e)

**Test Statistic and Critical Value**

- The Test Statistic (usual structure) is

$$TS = \frac{(-300) - 0}{105.9} = -2.83$$

- The Critical Value is determined with $\alpha = 0.05$, $k = 2$, $df = 16$.
- From statistical tables: $CV = 2.120$
- We can now apply the decision rule : Is the absolute value of the Test Statistic greater than the Critical Value?
- Is $2.83 > 2.12$? Yes We reject $H_0$. There is evidence of a difference in means.
- There is enough evidence to suggest that there is a difference in lifespans for the two brands.

# Example 4: Difference in Proportions (a)

- An experiment is conducted investigating the long-term effects of early childhood intervention programs (such as head start).
- In one experiment, the high-school drop out rate of the ***experimental group*** (which attended the early childhood program) and the ***control group*** (which did not) were compared.
- In the experimental group, 73 of 85 students graduated from high school.
- In the control group, only 43 of 82 students graduated. Is this difference statistically significant?
- (For this procedure, you may assume that the 0.05 level is chosen.)

# Example 4: Difference in Proportions (b)

- The first step in hypothesis testing is to specify the null hypothesis and an alternative hypothesis.
- When testing differences between proportions, the null hypothesis is that the two population proportions are equal.
- That is, the null hypothesis is:
  $H_0 : \pi_1 = \pi_2$
  $H_1 : \pi_1 \neq \pi_2$
- (Remark: Two Tailed Test k = 2, and $\alpha = 0.05$)

# Example 4: Difference in Proportions (c)

- The next step is to compute the difference between the sample proportions.
- In this example, The point estimat is $\hat{p}_1 - \hat{p}_2$

$$\hat{p}_1 - \hat{p}_2 = \frac{73}{85} - \frac{43}{82} = 0.8588 - 0.5244$$

.

- $\hat{p}_1 - \hat{p}_2 = 0.3344$.
- Difference in sample proportions is 33.44%

# Example 4: Difference in Proportions (d)

The formula for the estimated standard error is:

$$S.E(\hat{p}_1 - \hat{p}_2) = \sqrt{\bar{p}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

where $\bar{p}$ is a ***aggregate proportion*** (proportion of successes from overall sample, regardless of which group they are in).
(see next slide.)

# Example 4: Difference in Proportions (d)

**Aggregate Proportion**:

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} \times 100\% = \frac{73 + 43}{85 + 82} \times 100\% = \frac{116}{167} \times 100\% = 69.5\%$$

**Standard Error**:

$$S.E(\hat{p}_1 - \hat{p}_2) = \sqrt{69.5 \times 30.5 \left(\frac{1}{85} + \frac{1}{82}\right)} = 7.13\%$$

# Example 4: Difference in Proportions (e)

**Test Statistic**:

- Observed difference : 85.88% - 52.44% = 33.44%
- Calculation

$$i.e(73/85) - (43/82)$$

- Under the null hypothesis, the expected difference is zero.
- Test Statistic is therefore

$$T.S. = \frac{33.44\%}{7.13\%} = 4.69$$

# Example 4: Difference in Proportions (e)

- The Critical value is 1.96 (Large sample , $\alpha = 0.05$, k=2).
- The test statistic TS = 4.69, is greater than the critical value CV = 1.96, so we reject the null hypothesis.
- The conclusion is that the probability of graduating from high school is greater for students who have participated in the early childhood intervention program than for students who have not.