Lecture 10

The t-distribution. Two sample tests. Paired t-test.

We demonstrated earlier that linear combinations of Normal random variables are also Normally distributed. Thanks to the Central Limit Theorem $\bar{x}$ is approximately Normally distributed irrespective of the underlying distribution of X.

Thus the test statistic:

$$T = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

where μ is the hypothesised mean, n – the sample size, $\bar{x}$ is the sample mean and $\sigma = \sqrt{Var(X)}$.

In a practical context of a hypothesis test about the mean (unknown) of some measurement, how do we get information about σ?

The test procedure assumes that we know the value of σ.

In practice we have a sample from which we can get an estimate of σ:

$$\hat{\sigma} = \sqrt{\frac{s^2}{n-1}} = \sqrt{\frac{1}{n-1}\sum(x_i - \bar{x})^2}$$

and thus construct:

$$T = \frac{\bar{x} - \mu}{\frac{\hat{\sigma}}{\sqrt{n}}}$$

but this T will no longer have a N(0,1) distribution!

If n is large we can hope that $\hat{\sigma}$ is close to σ so that we can ignore the problem, but often the samples we deal with are small. If n<30 we need to make an adjustment.

The distribution is called a t-distribution. It has one parameter :

df=degrees of freedom = n-1

(= dimension of the space spanned by $x_i - \bar{x}$).

If $n = \infty$ the t is N(0,1). The formula for the pdf of t is not of major interest, the cdf and its inverse are available in Excel:

TDIST(x,df,tails)

x is the value at which we wish to evaluate the function
df is the degrees of freedom,
If *tails* =1 we get $1 - F_T(x)$, (p-value for one tailed tests)
If *tails* =2 we get $2(1 - F_T(x))$, (p-value for two tailed tests)
Any other value is an error.

The structure of the function reflects its use in hypothesis tests.

$TINV(\alpha, df)$

Returns the critical value c such that:

P(T<-c or T> c) = α.

Tables of t have been created – critical values for selected probabilities (α) and selected values of df. Linear interpolation is adequate for intermediate values.

Comparison of t and N(0,1):

x	t4	t30	N(0,1)
1	0.18695	0.162654	0.158655
2	0.058058	0.027313	0.02275
3	0.019971	0.002695	0.00135
4	0.008065	0.000191	3.17E-05

Critical values 2-tail:

p	t4	t30	N(0,1)
0.1	2.131846	1.697260	1.644853
0.05	2.776451	2.042270	1.959961
0.025	3.495406	2.359566	2.241395
0.01	4.604080	2.749985	2.575835

Example:

To be suitable for a certain assembly process solder must have a mean melting point in excess of 200° C. Five samples were tested and the following melting points were obtained.

199.9 200.6 201.2 200.7 199.8

$\bar{x} = 200.44$, $\hat{\sigma} = 0.585662$

We want the melting point to be in excess of 200. We assume it isn't and "hope" to reject H_0.

Thus:
H_0 $\mu = 200$ vs H_1 $\mu > 200$

We reject H_0 if T is too large.

$$T = \frac{200.44 - 200}{\frac{0.585662}{\sqrt{5}}} = 1.679928$$

If we were to assume σ known = 0.585662, using Normal distribution we get a p-value of 0.046486 i.e. just significant at 5% (critical value 1.64)

As the variance (sd) is estimated we should use the t-distribution on 5-1=4 degrees of freedom. This gives p-value=0.084135 , suspicious but not significant. So we do not conclude that mean melting point is above 200.

Two sample tests.

Two speaker independent voice recognition systems were tested repeatedly . The percentage of errors was recorded as follows:

Sys1	Sys2
7.98	7.80
8.1	7.92
7.15	6.26
8.8	8.35
8.78	8.03
9.37	9.01
8.04	
9.35	
9.86	
8.42	
9.58	
8.01	

The summary statistics are:

System 1 n_1=12, $\bar{x}$=8.62, $\hat{\sigma}$=0.809152
System 2 n_2=6, $\bar{x}$=7.895, $\hat{\sigma}$=0.911192

Is there evidence that one system is better than the other?

H_0: $\mu_1 = \mu_2$ H_1: $\mu_1 \neq \mu_2$

A priori we had no reason to believe system 2 has lower error rate, so a two tailed test.

This is equivalent to:

H_0: $\mu_1 - \mu_2 = 0$ vs H_1: $\mu_1 - \mu_2 \neq 0$

The test statistic is:

$$T = \frac{\bar{x}_1 - \bar{x}_2}{sd(\bar{x}_1 - \bar{x}_2)}$$

we shall reject H_0 if T is too far away from 0 in either direction.

$$sd(\bar{x}_1 - \bar{x}_2) = \sqrt{V(\bar{x}_1 - \bar{x}_2)} = \sqrt{V(\bar{x}_1) + V(\bar{x}_2)}$$

$$= \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 0.439248$$

If we know the variances we can calculate this and T has N(0,1) distribution under H_0.

$$T = \frac{0.725}{0.439248} = 1.650548 \qquad \text{p-value} = 0.049$$

If we do not know the variances and substitute estimates the distribution is **not** a simple t-distribution.

To get a simple t distribution we must *assume a common variance for both samples.* We then use a pooled estimate of variance:

$$\hat{\sigma}^2 = \frac{\sum(x_{1,i} - \bar{x}_1)^2 + \sum(x_{2,i} - \bar{x}_2)^2}{n_1 + n_2 - 2} = 0.709584$$

In practice for large samples (n>30) we plug in the estimates and pretend these are known variances (version 1).
For small samples (at least one of n <30) we get a pooled estimate and use t- distribution.

In this case:

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$ as t on $n_1 - n_2 - 2$ df.

$$T = \frac{0.725}{\sqrt{0.709584\left(\frac{1}{12} + \frac{1}{6}\right)}} = 1.721337 \text{ as } t_{6+12-2}$$

p- value =0.10464

The first version is called a *two sample z-test.*
The second *a two sample t-test.*

In fact there is an approximation to the t with unequal variances. Its just a t- distribution with different df than above –this is employed automatically by MINITAB, but not part of the theory of this course.

Paired t-test

Two speaker independent voice recognition systems were tested repeatedly . The percentage of errors was recorded as follows:

Individual	Sys1	Sys2
John	7.98	7.80
Mary	8.10	7.92
Ian	7.15	6.26
Carol	8.80	8.35
Richard	8.78	8.03
Jane	9.37	9.01

Note the difference in the experimental design here: Each individual tests both systems. Only six individuals here (there were 18 in the previous case).

With this type of data we can compare systems within each individual.

Method construct :

D_i = individual(i) sys 1 – individual (i) sys2

H_0: $E(D_i) = 0$ vs H_1: $E(D_i) \neq 0$

$T = \frac{\overline{D}}{sd(\overline{D})}$ as t_{n-1} (n individuals) , reject H_0 if T too far away from 0.

Di

0.18
0.18
0.89
0.45
0.75
0.36

$\overline{D} = 0.468333 \qquad \hat{\sigma}_D = 0.295121$

$$T = \frac{0.46833}{\frac{0.295121}{\sqrt{6}}} = 3.887138$$

p-value = 0.011558.

<u>Models</u>

The two sample tests are across individuals, the paired test is within individuals.

$$\text{Measurement} = \underset{\text{Mean}}{\mu} + \underset{\text{system}}{S_i} + \underset{\textit{Person}}{P_j} + \underset{\textit{error}}{E_{ij}}$$

The Mean and system are *systematic effects.*

The Person and error are random.

In the two sample test the variance is composed of two parts the Person effect and error effect.

In the paired test the differencing cancels the person effect out.

$$D_i = 0 \quad + \quad (S_1 \; - \; S_2) \; + \; 0 \quad + \quad (E_{i1} - E_{i2})$$

mean diff in Sys $(P_i\text{-}P_i)$ error

This means that the variation is smaller. The test is more powerful = chances of type II error are less.
The paring though must be real. There must be a common (person) effect to the pairs that are differenced.

For ref
If the above paired was treated as two sample

	Sys 1	Sys 2
	7.98	7.8
	8.1	7.92
	7.15	6.26
	8.8	8.35
	8.78	8.03
	9.37	9.01
mean	8.363333	7.895
sd	0.783088	0.911192
pooled var	0.721748	
sd Diff	0.490492	
T-stat	0.954823	