# The Theory of Designed Experiments

2. Randomization

UNESP-Botucatu, August 2010

# Why do we randomize?

Two reasons:

- After we have allowed for any expected patterns of variation among units there will usually still be many ways of allocating treatments to units. Doing it randomly removes any subjective element from the allocation.
- It justifies the usual form of analysis.

Remember our basic model, $y_{i(r)} = \mu + t_r + e_i$. What further assumptions do we need to make?

# Example

In a psychological experiment, 8 subjects, selected at random from a large population, perform a task and are scored out of 20. It is noted that the 4 female subjects get the highest scores.

| | | | |
|---|---|---|---|
| $M_1$ | 14 | $F_5$ | 15 |
| $M_2$ | 13 | $F_6$ | 15 |
| $M_3$ | 12 | $F_7$ | 16 |
| $M_4$ | 10 | $F_8$ | 19 |

Could this be coincidence or does it indicate that women perform this task better than men?

Interpretation is difficult - there are many possible coincidences. If we look at enough ways of splitting up the eight subjects we are bound to find some coincidence.

Before the experiment, what model would we have proposed?

# Example

4 subjects are selected at random from a large population who have been trained to perform the task and 4 are randomly selected from a large population who have not been trained. The 4 from the trained population get the highest scores.

| | | | |
|---|---|---|---|
| $C_1$ | 14 | $T_5$ | 15 |
| $C_2$ | 13 | $T_6$ | 15 |
| $C_3$ | 12 | $T_7$ | 16 |
| $C_4$ | 10 | $T_8$ | 19 |

What does this tell us about the effect of training?

There is fairly strong evidence of a difference between the populations. A two sample t-test gives a p-value of 0.026.

What would our model be in this case? Is it plausible?

# Example

8 employees of a company work in an analytical laboratory. 4 of the 8 are randomly selected to be trained to perform a new task. All subjects are then scored out of 20. The 4 who received training get the highest scores.

| | | | |
|---|---|---|---|
| $C_1$ | 14 | $T_5$ | 15 |
| $C_2$ | 13 | $T_6$ | 15 |
| $C_3$ | 12 | $T_7$ | 16 |
| $C_4$ | 10 | $T_8$ | 19 |

Does this indicate that training leads to better performance?

The alternative is that the randomization happened to select the 4 best subjects for training. We can quantify the probability that the randomization would have done this *without making any further assumptions*.

What is our model in this case?

# Randomization

This is the most important reason for randomizing treatments to units. In the first example the lack of randomization made it impossible to draw any such conclusions. In the second example conclusions were weaker and could only be drawn by making strong assumptions.

# Randomization

The above is an example of a *permutation test defined by the randomization*. An alternative (more general) way of carrying out the same test is as follows:

1. Consider all possible randomizations of treatments to units which could have occurred.
2. Assume the response in each unit would have remained unchanged, i.e. assume the treatments have no effect (assume $H_0$ is true).
3. For each randomization, calculate the variance ratio ($F$-ratio).
4. The p-value is the proportion of randomizations which would have given as large a variance ratio as the actual data.

The first stage in the test shows that the precise form of analysis is defined by how the experiment was randomized.

# Randomization

We will set up the theory for a completely randomized design for $n$ units with $t$ treatments each applied to $n_t = \frac{n}{t}$ units.

Express the model as

$$y_{i(r)} = \mu + t_r + e_i,$$

$\sum_i e_i = 0$ and $\sum_r t_r = 0$.

Perform randomization by:

1. writing down the *combinatorial design*;
2. randomly allocating units to unit labels.

# Randomization

Over the population of possible randomizations, the above deterministic model becomes the stochastic model.

$$Y_{i(r)} = \mu + t_r + \sum_{j=1}^{n} \delta_{ij} e_j,$$

where $\delta_{ij} = 1$ if unit $j$ is randomized to unit label $i$ and $\delta_{ij} = 0$ otherwise.

$\sum_{j=1}^{n} e_j = 0$ and $\sum_{j=1}^{n} e_j^2 = (n-1)\sigma^2$, where $\sigma^2$ is the variance of the $e_j$.

Then

$$\sum_{j=1}^{n} \sum_{\substack{l=1 \\ l \neq j}}^{n} e_j e_l = -\sum_{j=1}^{n} e_j^2 = -(n-1)\sigma^2.$$

# Randomization

We also have

$$
\begin{aligned}
E(\delta_{ij}) &= \frac{1}{n}; \\
E(\delta_{ij}^2) &= \frac{1}{n} \quad \Rightarrow \quad Var(\delta_{ij}) = \frac{n-1}{n^2}; \\
E(\delta_{ij}\delta_{il}) &= 0 \quad \Rightarrow \quad Cov(\delta_{ij}, \delta_{il}) = -\frac{1}{n^2}; \\
E(\delta_{ij}\delta_{kj}) &= 0 \quad \Rightarrow \quad Cov(\delta_{ij}, \delta_{kj}) = -\frac{1}{n^2}; \\
E(\delta_{ij}\delta_{kl}) &= \frac{1}{n(n-1)} \quad \Rightarrow \quad Cov(\delta_{ij}, \delta_{kl}) = \frac{1}{n^2(n-1)}.
\end{aligned}
$$

# Randomization

Letting $\epsilon_i = \sum_{j=1}^{n} \delta_{ij} e_j$, the model is

$$Y_{i(r)} = \mu + t_r + \epsilon_i.$$

From the previous results, it is easy to show that $E(\boldsymbol{\epsilon}) = 0$ and

$$V(\boldsymbol{\epsilon}) = \left( \mathbf{I} - \frac{1}{n} \mathbf{J} \right) \sigma^2,$$

where $\mathbf{J}$ is a matrix of 1s.

$\boldsymbol{\epsilon}$ satisfies the Gauss-Markov conditions, so least squares provides minimum variance linear unbiased estimators of all parameters and the residual mean square is the minimum variance unbiased estimator of $\sigma^2$.

# Randomization

Therefore the usual linear models analysis is still appropriate and is completely defined by the randomization which has been done.

No assumption has been made that the units are a random sample from some infinite population of potential units.

The Normal distributional results no longer hold, but are usually a reasonable approximation. If we do not trust them, we should use permutation tests and related confidence intervals.

# Randomization

If the experimental units *are* a random sample from a large population, the above theory still holds.

We might be able to make an additional assumption about the distribution of $\epsilon$, e.g. that it is Normal, to justify parametric inferences.

However, randomization is still useful because it provides additional robustness to the failure of our assumption, i.e. the linear model is still justified even if our assumption about the population is wrong.

# Randomization

Infinite Normal theory model:

- Assumes units are a random sample from a normally distributed population, with constant variance.
- Inferences apply to the infinite population from which the units were sampled.

Finite randomization theory model:

- Assumes a finite set of units to which treatments are allocated at random.
- Inferences apply only to the finite set of units.

So the analysis is always the same, but the conclusions which can be drawn from the results depend on how the experimental units were selected.

# Randomization

If randomization leads to a valid analysis, why do we ever restrict the randomization, e.g. by blocking?

- The analysis still depends on $\sigma^2$ and so may be inefficient.
- The randomization theory, with appropriate modification, follows for randomized block and other designs.