STAT 2011
Workshop on Data Exploration and Technical Writing

Section 3: ANOVA and Experiment

Reference: Dean A, Voss D, Draguljić D. Design and analysis of experiments 2nd edition. New York: Springer, 2007.

Dr. OUYANG Ming

2023/2024 Term 2

Reasons for an Experiment

- To determine the principal causes of variation in a measured response.
- To find the conditions that give rise to a maximum or minimum response.
- To compare the responses achieved at different settings of controllable variables.
- To obtain a mathematical model in order to predict future response.

Observations can be collected from observational studies as well as from experiments, but **only an experiment allows conclusions to be drawn about cause and effect.**

Reasons for an Experiment

Primary
Data Collection

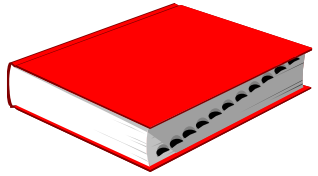
- **Observation**

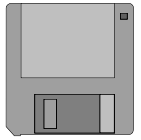
- **Survey**

- **Experimentation**

Secondary
Data Compilation

- **Print or Electronic**





Quality? Accuracy?

How to get the data?

Meet the assumptions?

Reasons for an Experiment

The output from each machine on a factory floor is constantly monitored by any successful manufacturing company. Suppose that in a particular factory, the output from a particular machine is consistently of low quality.

What should the managers do?

- ✓ They could conclude that the machine needs replacing and pay out a large sum of money for a new one.
- ✓ They could decide that a machine operator is at fault and dismiss him/her.
- ✓ They could conclude that the humidity in that part of the factory is too high and install a new air conditioning system.
- ✓ ……

Reasons for an Experiment

Although it has been very effective in showing the management that a problem exists, it has given them **very little idea** about the **cause** of the poor quality.

It would actually be a simple matter to determine or rule out some of the potential causes.

This is an experiment.

- ✓ The experimenter has control over a possible cause in the difference in output quality between machines.
- ✓ If this particular cause is ruled out, then the experimenter can begin to vary other factors.
- ✓ It is more efficient to examine all possible causes of variation simultaneously rather than one at a time.

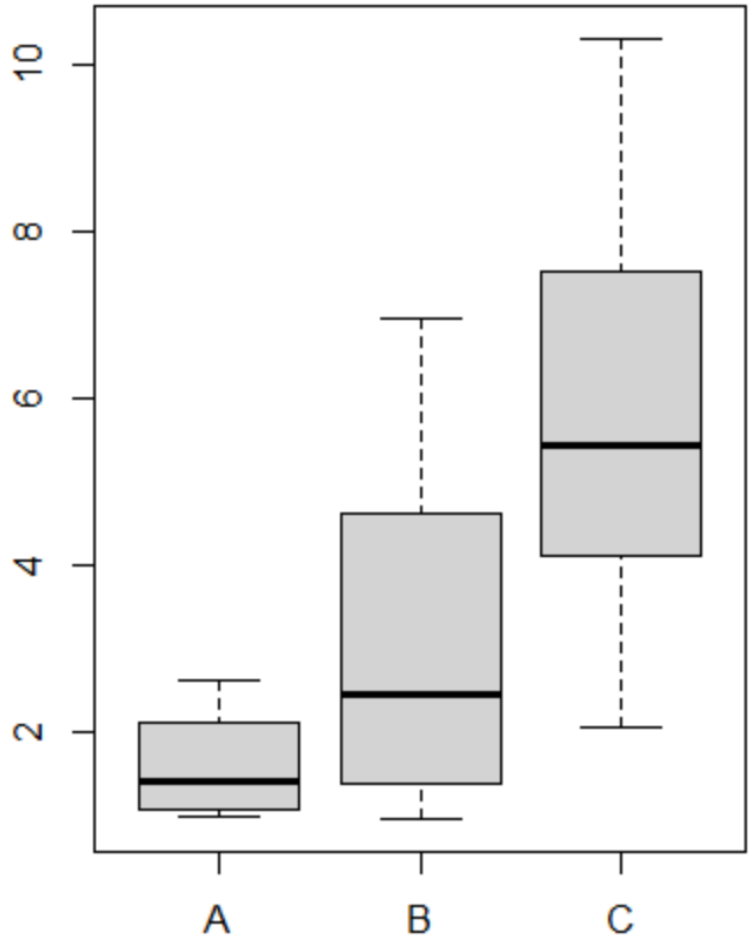
Reasons for an Experiment

Data from Particular factory

Machines

Rate of low quality

	1	2	3	4	5	6	7	8
Low humidity	1.00	1.01	1.13	1.14	1.70	2.01	2.23	2.63
Moderate humidity	0.96	1.23	1.54	1.96	2.94	3.68	5.59	6.96
High humidity	2.07	3.72	4.50	4.90	6.00	6.84	8.23	10.33



Hypothesis testing

Two sample T-test

σ_1^2, σ_2^2 are Unknown $\Rightarrow \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}$ are Unknown

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2)/n + (\sigma_2^2)/m}} \sim Normal(0,1)$$

Estimate population variance by sample variance

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{(S_1^2)/n + (S_2^2)/m}}$$

Assumptions:

- Normal distribution is required (Linear combination):
  $X_1, \cdots, X_n \sim Normal(\mu_1, \sigma_1^2)$
  $Y_1, \cdots, Y_m \sim Normal(\mu_2, \sigma_2^2)$
- If σ_1^2, σ_2^2 are Unknown but **equal to each other:** $\sigma_1^2 = \sigma_2^2 = \sigma^2$

Define the **pooled estimate of variance** as:

$$S_p^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n + m - 2}$$

which is a weighted average of two sample variances.

$$E(S_p^2) = \sigma^2 = \sigma_1^2 = \sigma_2^2$$

Hypothesis testing

Two sample T-test

σ_1^2, σ_2^2 are Unknown $\Rightarrow \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}$ are Unknown

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2)/n + (\sigma_2^2)/m}} \sim Normal(0,1)$$

Estimate population variance by sample variance

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{(S_p^2)/n + (S_p^2)/m}}$$

Assumptions:

- Normal distribution is required (Linear combination):

  $X_1, \cdots, X_n \sim Normal(\mu_1, \sigma_1^2)$

  $Y_1, \cdots, Y_m \sim Normal(\mu_2, \sigma_2^2)$

- If σ_1^2, σ_2^2 are Unknown but **equal to each other:** $\sigma_1^2 = \sigma_2^2 = \sigma^2$

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_p\sqrt{1/n + 1/m}} \qquad t - distribution\ (n + m - 2)$$

$$= \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sigma\sqrt{1/n + 1/m}} \div \sqrt{\left[\frac{(n-1)S_1^2}{\sigma^2} + \frac{(m-1)S_2^2}{\sigma^2}\right]/(n + m - 2)}$$

Hypothesis testing

Two sample T-test

σ_1^2, σ_2^2 are Unknown $\Rightarrow \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}$ are Unknown

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2)/n + (\sigma_2^2)/m}} \sim Normal(0,1)$$

Estimate population variance by sample variance

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{(S_p^2)/n + (S_p^2)/m}}$$

$$t_0 = \frac{\bar{x} - \bar{y}}{s_p\sqrt{1/n + 1/m}}$$

$$\begin{cases} H_0: \mu_1 - \mu_2 = 0 \\ H_1: \mu_1 - \mu_2 \neq 0 \end{cases}$$

Reject H_0 if $|t_0| > t_{n+m-2,1-\alpha/2}$, then H_1 is accept

$$\begin{cases} H_0: \mu_1 - \mu_2 = 0 \\ H_1: \mu_1 - \mu_2 > 0 \end{cases}$$

Reject H_0 if $t_0 > t_{n+m-2,1-\alpha}$, then H_1 is accept

$$\begin{cases} H_0: \mu_1 - \mu_2 = 0 \\ H_1: \mu_1 - \mu_2 < 0 \end{cases}$$

Reject H_0 if $t_0 < t_{n+m-2,\alpha}$, then H_1 is accept

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

Multiple hypothesis testing

$$\begin{cases} H_0: \mu_1 - \mu_2 = 0 \\ H_1: \mu_1 - \mu_2 \neq 0 \end{cases}$$

$$\begin{cases} H_0: \mu_1 = \mu_2 = \mu_3 \\ H_1: at\ least\ two\ of\ them\ differ \end{cases}$$

$$\begin{cases} H_0: \mu_1 - \mu_3 = 0 \\ H_1: \mu_1 - \mu_3 \neq 0 \end{cases} \qquad \begin{cases} H_0: \mu_2 - \mu_3 = 0 \\ H_1: \mu_2 - \mu_3 \neq 0 \end{cases}$$

Statistical hypothesis testing is based on **rejecting the null hypothesis** if the probability of observed data under the null hypotheses is low. **If multiple hypotheses are tested, the chance of observing a rare event increases**, and therefore, the probability of incorrectly rejecting a null hypothesis (i.e., making a Type I error) increases.

$$\Pr\left\{\bigcup_{i=1}^{3}(p - value_i < \alpha)\right\} \leq \sum_{i=1}^{3} \Pr\{p - value_i < \alpha\} = 3\alpha$$

Bonferroni correction rejects the null hypothesis for each $p - value_i \leq \frac{\alpha}{m}$

Analysis of Variance

One-way ANOVA

Means of Several Independent Sample

$$Y_{it} = \mu + \tau_i + \epsilon_{it},$$
$$\epsilon_{it} \sim N(0, \sigma^2),$$
ϵ_{it}'s are mutually independent,
$$t = 1, \ldots, r_i, \quad i = 1, \ldots, v,$$

$$\begin{cases} H_0: \mu_1 = \mu_2 = \cdots = \mu_v \\ H_1: at\ least\ two\ of\ them\ differ \end{cases}$$

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Ratio	Expected Mean Square
Treatments	$v-1$	SST	$SST/(v-1)$	MST/MSE	$\sigma^2 + Q(\tau_i)$
Error	$n-v$	SSE	$SSE/(n-v)$		σ^2
Total	$n-1$	$SSTotal$			

Computational Formulae

$SST = \sum_i r_i(\bar{y}_{i.} - \bar{y}_{..})^2$ $\qquad$ $SSE = \sum_i \sum_t (y_{it} - \bar{y}_{i.})^2$

$SSTotal = SST + SSE = \sum_i \sum_t (y_{it} - \bar{y}_{..})^2$

$Q(\tau_i) = \sum_i r_i(\tau_i - \sum_h r_h \tau_h / n)^2/(v-1)$

Reject H_0 if $MST/MSE > F_{v-1,n-v,\alpha}$.

Analysis of Variance

$$X_{it} = \mu + \tau_i + \epsilon_{it}$$

$$\boldsymbol{\epsilon_{it} \sim N(0, \sigma^2)}$$

$$i = 1,2,3\ (Treatment)$$

$$t = 1, \cdots, n_i (Sample\ sizes)$$

Normal assumption &
Variance equal assumption

$$\begin{cases} H_0: \tau_1 = \tau_2 = \tau_3 \\ H_1: at\ least\ two\ of\ them\ differ \end{cases}$$

$\overline{x_{1\cdot}}, \overline{x_{2\cdot}}, \overline{x_{3\cdot}}, \overline{x_{\cdot\cdot}}$ should also follows normal distribution

One-way ANOVA

Means of Several Independent Sample

Treatment	A	B	C	
1	x_{11}	x_{21}	x_{31}	
2	x_{12}	x_{22}	x_{32}	
3	x_{13}	x_{23}	x_{33}	
4	x_{14}	x_{24}	x_{34}	
5	x_{15}	x_{25}	x_{35}	
	$\cdots/n_1$	$\cdots/n_2$	$\cdots/n_3$	
Mean	$\overline{x_{1\cdot}}$	$\overline{x_{2\cdot}}$	$\overline{x_{3\cdot}}$	$\overline{x_{\cdot\cdot}}$

$$\frac{x_{11} + \cdots + x_{1n_1} + x_{21} + \cdots + x_{2n_2} + x_{31} + \cdots + x_{3n_3}}{n_1 + n_2 + n_3} = \frac{n_1\overline{x_{1\cdot}} + n_2\overline{x_{2\cdot}} + n_3\overline{x_{3\cdot}}}{n_1 + n_2 + n_3}$$

Analysis of Variance

$$SSTotal = \sum_{i=1}^{3}\sum_{t=1}^{n_i}(x_{it} - \overline{x_{..}})^2$$

$$= \sum_{i=1}^{3}\sum_{t=1}^{n_i}(x_{it} - \overline{x_{i.}} + \overline{x_{i.}} - \overline{x_{..}})^2$$

$$= \sum_{i=1}^{3}\sum_{t=1}^{n_i}(x_{it} - \overline{x_{i.}})^2 + \sum_{i=1}^{3}\sum_{t=1}^{n_i}(\overline{x_{i.}} - \overline{x_{..}})^2$$

$$= SSE + SST$$

$$for\ i = 1,2,3, \sum_{t=1}^{n_i}(x_{it} - \overline{x_{i.}})(\overline{x_{i.}} - \overline{x_{..}}) = 0$$

One-way ANOVA

Means of Several Independent Sample

Treatment	A	B	C	
1	x_{11}	x_{21}	x_{31}	
2	x_{12}	x_{22}	x_{32}	
3	x_{13}	x_{23}	x_{33}	
4	x_{14}	x_{24}	x_{34}	
5	x_{15}	x_{25}	x_{35}	
	$\cdots/n_1$	$\cdots/n_2$	$\cdots/n_3$	
Mean	$\overline{x_{1.}}$	$\overline{x_{2.}}$	$\overline{x_{3.}}$	$\overline{x_{..}}$

Analysis of Variance

$$SST = \sum_{i=1}^{3}\sum_{t=1}^{n_i}(\overline{x_{i.}} - \overline{x_{..}})^2$$

$$f_0 = \frac{SST/(3-1)}{SSE/(n_1+n_2+n_3-3)}$$

$$SSE = \sum_{i=1}^{3}\sum_{t=1}^{n_i}(x_{it} - \overline{x_{i.}})^2$$

F-distribution

Reject H_0 if $f_0 > F_{3-1,n_1+n_2+n_3-3-1,1-\alpha}$ or then H_1 is accept

One-way ANOVA

Means of Several Independent Sample

Treatment	A	B	C	
1	x_{11}	x_{21}	x_{31}	
2	x_{12}	x_{22}	x_{32}	
3	x_{13}	x_{23}	x_{33}	
4	x_{14}	x_{24}	x_{34}	
5	x_{15}	x_{25}	x_{35}	
	$\cdots/n_1$	$\cdots/n_2$	$\cdots/n_3$	
Mean	$\overline{x_{1.}}$	$\overline{x_{2.}}$	$\overline{x_{3.}}$	$\overline{x_{..}}$

Analysis of Variance

$$SSTotal = \sum_{i=1}^{v}\sum_{j=1}^{u}\left(x_{ij} - \overline{x_{..}}\right)^2$$

$$X_{ij} = \mu + \tau_i + \gamma_j + \epsilon_{ij}$$

$$\boldsymbol{\epsilon_{ij} \sim N(0, \sigma^2)}$$

$i = 1,2,3,\cdots,v\ (Treatment\ A)$

$j = 1,2,3,\cdots,u\ (Treatment\ B)$

TWO-way ANOVA

Without Replacement

Treatment	A	B	C	
T1	x_{11}	x_{21}	x_{31}	$\overline{x_{.1}}$
T2	x_{12}	x_{22}	x_{32}	$\overline{x_{.2}}$
T3	x_{13}	x_{23}	x_{33}	$\overline{x_{.3}}$
T4	x_{14}	x_{24}	x_{34}	$\overline{x_{.4}}$
T5	x_{15}	x_{25}	x_{35}	$\overline{x_{.5}}$
Mean	$\overline{x_{1.}}$	$\overline{x_{2.}}$	$\overline{x_{3.}}$	$\overline{x_{..}}$

$$\begin{cases} H_{0A}: \tau_1 = \tau_2 = \cdots = \tau_v = 0 \\ H_{0B}: \gamma_1 = \gamma_2 = \cdots = \gamma_u = 0 \\ H_1: at\ least\ one\ of\ them\ \tau_i\ or\ \gamma_j\ not\ equal\ to\ 0 \end{cases}$$

Analysis of Variance

TWO-way ANOVA

Without Replacement

$$SSTotal = \sum_{i=1}^{v}\sum_{j=1}^{u}\left(x_{ij} - \overline{x_{..}}\right)^2 = \sum_{i=1}^{v}\sum_{j=1}^{u}\left(x_{ij} + (\overline{x_{i.}} - \overline{x_{..}}) + \left(\overline{x_{.j}} - \overline{x_{..}}\right) - \overline{x_{i.}} - \overline{x_{.j}} + \overline{x_{..}}\right)^2$$

$$X_{ij} = \mu + (\mu_i - \mu) + (\mu_j - \mu) + \epsilon_{ij}$$

$$\boldsymbol{\epsilon_{ij} \sim N(0, \sigma^2)}$$

$$i = 1,2,3,\cdots,v\ (Treatment\ A)$$

$$\mathrm{j} = 1,2,3,\cdots,u\ (Treatment\ B)$$

$$= v\sum_{j=1}^{u}(\overline{x_{i.}} - \overline{x_{..}})^2 + u\sum_{i=1}^{v}\left(\overline{x_{.j}} - \overline{x_{..}}\right)^2 + \sum_{i=1}^{v}\sum_{j=1}^{u}\left(x_{ij} - \overline{x_{i.}} - \overline{x_{.j}} + \overline{x_{..}}\right)^2$$

$$= SSTA + SSTB + SSE$$

$$\begin{cases} H_{0A}: \tau_1 = \tau_2 = \cdots = \tau_v = 0 \\ H_{0B}: \gamma_1 = \gamma_2 = \cdots = \gamma_u = 0 \\ H_1: at\ least\ one\ of\ them\ \tau_i\ or\ \gamma_j\ not\ equal\ to\ 0 \end{cases}$$

Analysis of Variance

$$SSTotal = \sum_{i=1}^{v}\sum_{j=1}^{u}\left(x_{ij} - \overline{x_{..}}\right)^2$$

$$= SSTA + SSTB + SSE$$

$$\begin{cases} H_{0A}: \tau_1 = \tau_2 = \cdots = \tau_v = 0 \\ H_{0B}: \gamma_1 = \gamma_2 = \cdots = \gamma_u = 0 \\ H_1: at\ least\ one\ of\ them\ \tau_i\ or\ \gamma_j\ not\ equal\ to\ 0 \end{cases}$$

TWO-way ANOVA

Without Replacement

F-distribution

$$f_A = \frac{SSTA/(v-1)}{SSE/[(u-1)(v-1)]}$$

Reject H_{0A} if $f_A > F_{v-1,(u-1)(v-1),1-\alpha}$ or then H_1 is accept

$$f_B = \frac{SSTB/(u-1)}{SSE/[(u-1)(v-1)]}$$

Reject H_{0B} if $f_B > F_{u-1,(u-1)(v-1),1-\alpha}$ or then H_1 is accept

Analysis of Variance

TWO-way ANOVA

With Replacement

		Relaxation exercise: muscle relaxation	Relaxation exercise: guided imagery	Total
Physical therapy	stretching	6 participants (cell a)	6 participants (cell b)	12
	strengthening	6 participants (cell c)	6 participants (cell d)	12
	Total	12	12	24

Hypothesis 1. Test of Physical Therapy: Stretching vs. Strengthening
Hypothesis 2. Test of Relaxation Exercise: Muscle Relaxation vs. Guided Imagery
Hypothesis 3. Test of the **Interaction Effect** of Physical Therapy and Relaxation Exercise

Analysis of Variance

$$X_{ijt} = \mu + \tau_i + \gamma_j + (\tau\gamma)_{ij} + \epsilon_{ijt}$$

$$\epsilon_{ijt} \sim N(0, \sigma^2)$$

$$i = 1,2,3,\cdots,v\ (Treatment\ A)$$

$$\mathrm{j} = 1,2,3,\cdots,u\ (Treatment\ B)$$

$$t = 1,\cdots,n_{ij}(Sample\ sizes)$$

$$SSTotal = \sum_{i=1}^{v}\sum_{j=1}^{u}\sum_{t=1}^{n_{ij}}\left(x_{ijt} - \overline{x_{...}}\right)^2$$

TWO-way ANOVA

With Replacement

$$\begin{cases} H_{0A}: \tau_1 = \tau_2 = \cdots = \tau_v = 0 \\ H_{0B}: \gamma_1 = \gamma_2 = \cdots = \gamma_u = 0 \\ H_{0AB}: \tau\gamma_{11} = \tau\gamma_{12} = \cdots = \tau\gamma_{vu} = 0 \\ H_1: at\ least\ one\ of\ them\ not\ equal\ to\ 0 \end{cases}$$

$$SSTA = \sum_{i=1}^{v}\sum_{j=1}^{u} n_{ij}\left(\overline{x_{i..}}^{A} - \overline{x_{...}}\right)^2$$

Main effect

$$SSTB = \sum_{j=1}^{u}\sum_{i=1}^{v} n_{ij}\left(\overline{x_{.j.}}^{B} - \overline{x_{...}}\right)^2$$

Analysis of Variance

TWO-way ANOVA

With Replacement

$$X_{ijt} = \mu + \tau_i + \gamma_j + (\tau\gamma)_{ij} + \epsilon_{ijt}$$

$$\epsilon_{ijt} \sim N(0, \sigma^2)$$

$$i = 1,2,3,\cdots,v\ (Treatment\ A)$$

$$\mathrm{j} = 1,2,3,\cdots,u\ (Treatment\ B)$$

$$t = 1,\cdots,n_{ij}(Sample\ sizes)$$

$$SSE = \sum_{i=1}^{v}\sum_{j=1}^{u}\sum_{t=1}^{n_{ij}}\left(x_{ijt} - \overline{x_{ij.}}^{AB}\right)^2$$

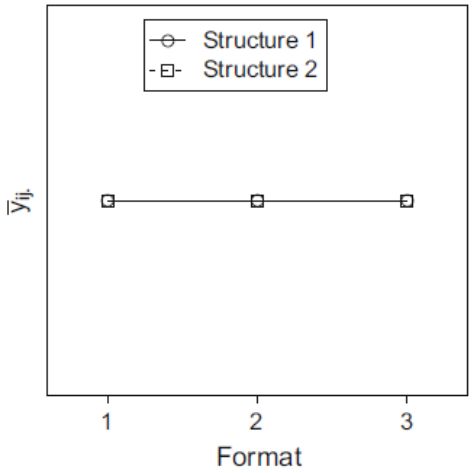
$$SSTAB = \sum_{i=1}^{v}\sum_{j=1}^{u}\sum_{t=1}^{n_{ij}}\left(\overline{x_{ij.}}^{AB} - \overline{x_{i..}}^{A} - \overline{x_{.j.}}^{B} + \overline{x_{...}}\right)^2$$

$$SSTotal = \sum_{i=1}^{v}\sum_{j=1}^{u}\sum_{t=1}^{n_{ij}}\left(x_{ijt} - \overline{x_{...}}\right)^2 = SSTA + SSTB + SSE + \boxed{SSTAB}$$

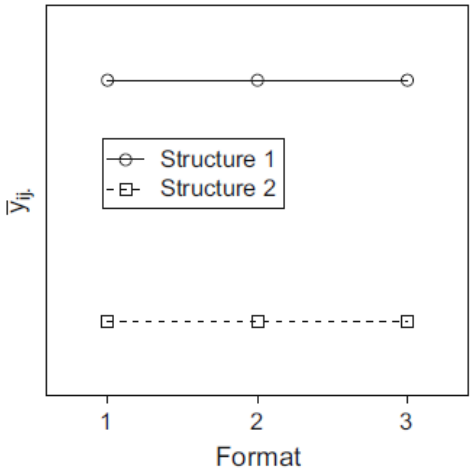
Interaction effect

Analysis of Variance

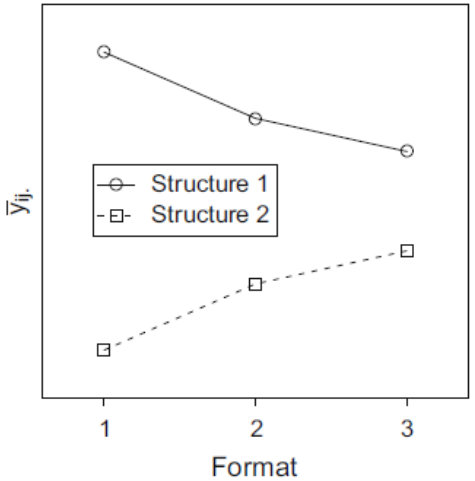
Interaction effect



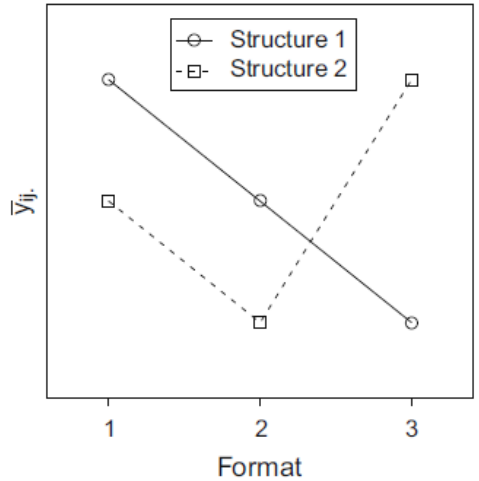
(a) F=no, S=no, FS=no



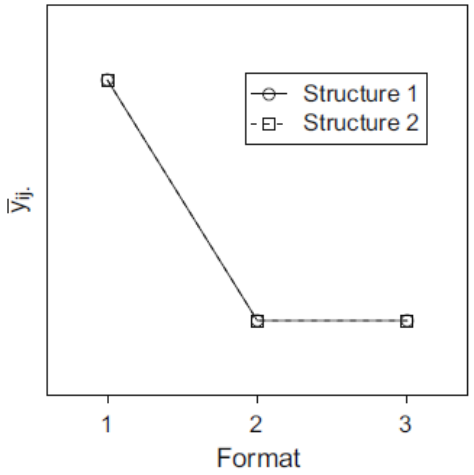
(b) F=no, S=yes, FS=no



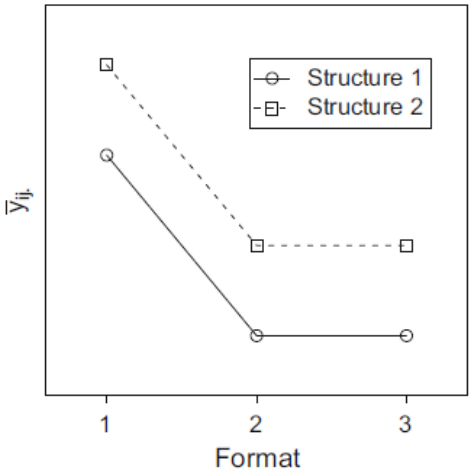
(a) F=no?, S=yes, FS=yes



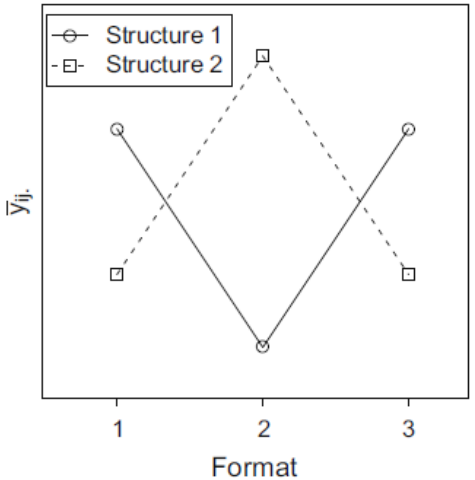
(b) F=yes, S=no?, FS=yes



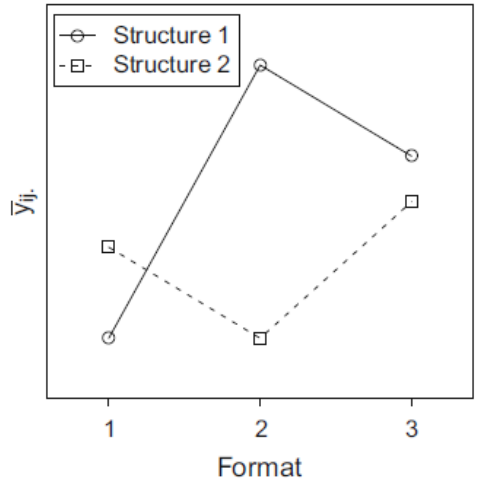
(c) F=yes, S=no, FS=no



(d) F=yes, S=yes, FS=no



(c) F=no?, S=no?, FS=yes



(d) F=yes, S=yes, FS=yes

Analysis of Variance

$$f_A = \frac{SSTA/(v-1)}{SSE/(\sum_{i=1}^{v}\sum_{j=1}^{u} n_{ij} - uv)}$$

Reject H_{0A} if $f_A > F_{v-1,\sum_{i=1}^{v}\sum_{j=1}^{u} n_{ij}-uv,1-\alpha}$ or then H_1 is accept

$$f_B = \frac{SSTB/(u-1)}{SSE/(\sum_{i=1}^{v}\sum_{j=1}^{u} n_{ij} - uv)}$$

Reject H_{0B} if $f_B > F_{u-1,\sum_{i=1}^{v}\sum_{j=1}^{u} n_{ij}-uv,1-\alpha}$ or then H_1 is accept

TWO-way ANOVA

With Replacement

$$\begin{cases} H_{0A}: \tau_1 = \tau_2 = \cdots = \tau_v = 0 \\ H_{0B}: \gamma_1 = \gamma_2 = \cdots = \gamma_u = 0 \\ H_{0AB}: \tau\gamma_{11} = \tau\gamma_{12} = \cdots = \tau\gamma_{vu} = 0 \\ H_1: at\ least\ one\ of\ them\ not\ equal\ to\ 0 \end{cases}$$

$$f_{AB} = \frac{SSTAB/((v-1)(u-1))}{SSE/(\sum_{i=1}^{v}\sum_{j=1}^{u} n_{ij} - uv)}$$

Reject H_{0AB} if $f_{AB} > F_{v-1,\sum_{i=1}^{v}\sum_{j=1}^{u} n_{ij}-uv,1-\alpha}$ or then H_1 is accept

The Art of Designing an Experiment

- The art of **designing an experiment** and the art of **analyzing an experiment** are closely intertwined and need to be studied side by side.
- In designing and experiment, one must take into account the analysis that will be performed.
- In turn, the efficiency of the analysis will depend upon the particular experimental design that is used to collect data.

Without these considerations, it is possible to invest much time, effort, and expense in the collection of data which in fact contribute little to the research questions.

A guiding principle of experimental design is to keep it simple.

(Interpretation and presentation of the results of experiments are generally clearer for simpler experiments.)

Exploratory and Confirmatory

The purpose of an experiment can range from **exploratory** (discovering new important sources of variability) to **confirmatory** (confirming that previously discovered sources of variability are sufficiently major to warrant further study), and the philosophy of the analysis depends on the purpose of the experiment.

In the early stages of experimentation the analysis may be exploratory, and one would plot and analyze the data in any way that assists in the identification of important sources of variation.

In later stages of experimentation, analysis is usually confirmatory in nature.

Exploratory and Confirmatory

A mathematical model of the response is postulated and hypotheses are tested and confidence intervals are calculated.

Our models include **random error variables** that encompass all the sources of variability not explicitly present in the model. **[Independent]**

Replication ←

There is a difference between **replication** and **repeated measurements**.

Suppose four subjects are each assigned to a drug and a measurement is taken on each subject. The result is four independent observations on a drug.

If one subject is assigned to a drug and then measured four times, the measurement are not independent.

Exploratory and Confirmatory

A mathematical model of the response is postulated and hypotheses are tested and confidence intervals are calculated.

Our models include **random error variables** that encompass all the sources of variability not explicitly present in the model. **[Independent]**

Replication ←

There is a difference between **replication** and **repeated measurements**.

The variation recorded in repeated measurements taken at the same time reflects the variation in the measurement process, taken over a time interval reflects the variation in the single subject's response to the drug over time. Neither reflect the variation in independent subject's response to the drug.

Exploratory and Confirmatory

A mathematical model of the response is postulated and hypotheses are tested and confidence intervals are calculated.

Our models include **random error variables** that encompass all the sources of variability not explicitly present in the model. **[Independent]**

Replication

There is a difference between **replication** and **repeated measurements**.

We need to know about the variation in independent subjects' responses to the drug so that we can generalize any conclusion about the drug so that it is relevant to all similar subjects.

A Checklist for Planning Experiments

The steps in the following checklist summarize a very large number of decisions that need to be made at each stage of the experimental planning process.
The steps are not independent, and at any stage, it may be necessary to go back and revise some of the decisions made at an earlier stage.

1. Define the objectives of the experiment.
2. Identify all sources of variation, including:
   A. treatment factors and their levels,
   B. experimental units,
   C. noise factors, and covariates.
3. Choose a rule for assigning the experimental units to the treatments.
4. Specify the measurements to be made, the experimental procedure, and the anticipated difficulties.

A Checklist for Planning Experiments

The steps in the following checklist summarize a very large number of decisions that need to be made at each stage of the experimental planning process.
The steps are not independent, and at any stage, it may be necessary to go back and revise some of the decisions made at an earlier stage.

5. Run a pilot experiment.
6. Specify the model.
7. Outline the analysis.
8. Review the above decisions. Revise, if necessary.

Define the objectives of the experiment.

- A list should be made of the precise questions that are to be addressed by the experiment.
- It is this list that helps to determine the decisions required at the subsequent stages of the checklist.
- It is advisable to list only the essential questions, since side issues will unnecessarily complicate the experiment, increasing both the cost and the likelihood of mistakes

In compiling the list of objects, it can often be helpful to outline the conclusions expected from the analysis of the data.

Identify all sources of variation: Treatment factors and their levels

- A source of variation is **anything** that could cause an observation to have a different numerical value from another observation.
- It is good practice to make a list of every conceivable source of variation and then label each as either major or minor.
- Major sources of variation can be divided into two types:
  - Those that are of particular interest to the experimenter, called **treatment factors**;
  - Those that are not of interest, called **nuisance factors**.

Levels should be selected: The levels are the specific types or amounts of the treatment factor that will actually be used in the experiment.

Treatment factors are often labeled F1, F2, F3, ... or A, B, C, ...

Identify all sources of variation: Treatment factors and their levels

- For example, a treatment factor might be a drug or a chemical additive or temperature or teaching method, etc.

- The levels of such treatment factors might be the different amounts of the drug to be studied, different types of chemical additives to be considered, selected temperature settings in the range of interest, different teaching methods to be compared, etc.

- If the levels of a treatment factor are quantitative, they are usually chosen to be equally spaced. For convenience, treatment factor levels can be coded. (Temperature levels 60 , 70 , 80 , ... might be coded as 1, 2, 3, ...)

Identify all sources of variation: Experimental Units

Experimental units are the **material** to which the levels of the treatment factor(s) are applied.

- For example, in agriculture these would be individual plots of land, in medicine they would be human or animal subjects, in industry they might be batches of raw material, factory workers, etc.

Experimental units should be representative of the material and conditions to which the conclusions of the experiment will be applied.

Choose a rule for assigning the experimental units to the treatments

A **completely randomized design** is the name given to a design in which the experimenter assigns the experimental units to the treatments complete at random, subject only to the number of observations to be taken on each treatment.

The purpose of randomization is **to prevent systematic and personal biases** from being introduced into the experiment by the experimenter.

A random assignment of subjects or experimental material to treatments prior to the start of the experiment ensures that observations that are favored or adversely selected by unknown sources of variation are observations selected in the luck of the draw and not systematically selected.

Choose a rule for assigning the experimental units to the treatments

A **completely randomized design** is the name given to a design in which the experimenter assigns the experimental units to the treatments complete at random, subject only to the number of observations to be taken on each treatment.

Consider an experiment to compare the effects on blood pressure of three exercise programs, where each program is observed four times, giving a total of 12 observations. Now, given 12 subjects, imagine making a list of all possible assignments of the 12 subjects to the three exercise programs so that 4 subjects are assigned to each program. (There are $12! = (4!\,4!\,4!) = 34650$ ways to do this.)
If the assignment of subjects to programs is done in such a way that **every possible assignment has the same chance of occurring**, then the assignment is said to be a completely random assignment.

It is, of course, possible that a random assignment itself could lead to the order 1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 3. An experimenter **SHOULD NOT** look at the resulting assignment, decide that it does not look very random, and change it.

Specify the Measurements to Be Made

The data (or observations) collected from an experiment are measurements of a response variable (e.g., the yield of a crop, the time taken for the occurrence of a chemical reaction, the output of a machine).

The units in which the measurements are to be made should be specified, and these should reflect the objectives of the experiment.

For example, if the experimenter is interested in detecting a difference of 0.5 gram in the response variable arising from two different treatments, it would not be sensible to take measurements to the nearest gram. On the other hand, it would be unnecessary to take measurements to the nearest 0.01 gram. Measurements to the nearest 0.1 gram would be sufficiently sensitive to detect the required difference, if it exists.

Specify the Experimental procedure and the Anticipated Difficulties

There are usually unforeseen difficulties in collecting data, but these can often be identified by taking a few practice measurements or by running a pilot experiment.

Listing the anticipated difficulties helps to identify sources of variation required by the checklist, and also gives the opportunity of simplifying the experimental procedure before the experiment begins

Run a Pilot Experiment

A pilot experiment is a mini experiment involving only a few observations. No conclusions are necessarily expected from such an experiment. It is run to aid the completion of the checklist.

- It provides an opportunity to practice the experimental technique and to identify unsuspected problems in the data collection.
- If the pilot experiment is large enough, it can also help in the selection of a suitable model for the main experiment.

Run a Pilot Experiment

Normal assumption??

Comparing Data Characteristics to Theoretical Properties

Do approximately **68%** of the observations lie within mean ± 1 standard deviation?

Do approximately **95%** of the observations lie within mean ± 2 standard deviation?

Do approximately **99%** of the observations lie within mean ± 3 standard deviation?

Is **the interquartile range approximately** 1.35σ?

Do stem-and-leaf display histogram or polygon appear bell-shaped **look symmetric**?

Do **the mean, median and mode** have **similar values**?

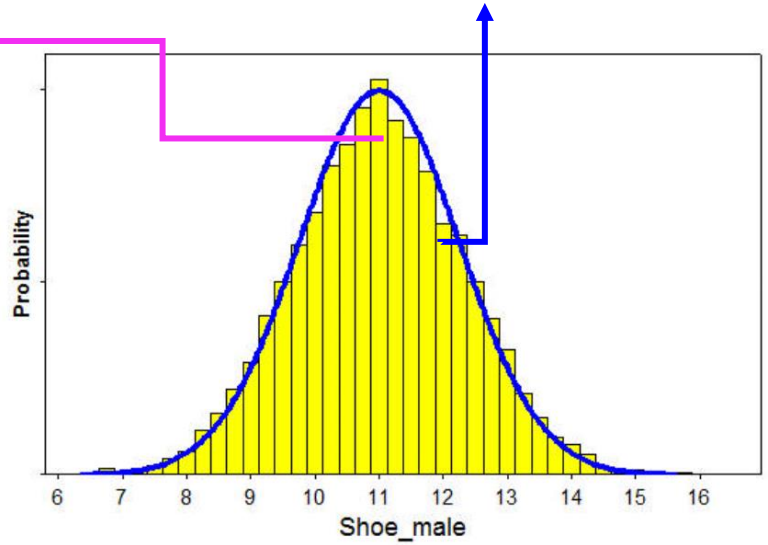
$\mu \pm 1\sigma$ covers about **68.26%** of X

$\mu \pm 2\sigma$ covers about **95.44%** of X

$\mu \pm 3\sigma$ covers about **99.73%** of X

For $Z \sim Normal(0,1); IQR = Q_3 - Q_1 \approx 1.35$

If $X \sim Normal(\mu, \sigma^2), IQR \approx 1.35\sigma$



Run a Pilot Experiment

Normal assumption??

Constructing the Normal Probability Plot / Quantile–quantile plot

Step1: Arrange data into ordered array (from smallest to the largest)

Step2: **Find corresponding standardized normal quantile values**

Step3: Plot the pairs of points with observed data values on the vertical axis and the standardized normal quantile values on the horizontal axis.

Step4: Evaluate the plot for evidence of linearity

For example, if you have a sample of $n = 19$, the $Z - value$ for the smallest value corresponds to a cumulative area of

$$\frac{1}{n+1} = \frac{1}{19+1} = \frac{1}{20} = 0.05$$

The 5% quantile of standard normal is -1.65.

Ordered Value	Z Value	Ordered Value	Z Value	Ordered Value	Z Value
1	−1.65	8	−0.25	14	0.52
2	−1.28	9	−0.13	15	0.67
3	−1.04	10	−0.00	16	0.84
4	−0.84	11	0.13	17	1.04
5	−0.67	12	0.25	18	1.28
6	−0.52	13	0.39	19	1.65
7	−0.39				

95% quantile

$0.05 * 4 = 20\%$ quantile of standard normal distribution

Run a Pilot Experiment

Normal assumption??

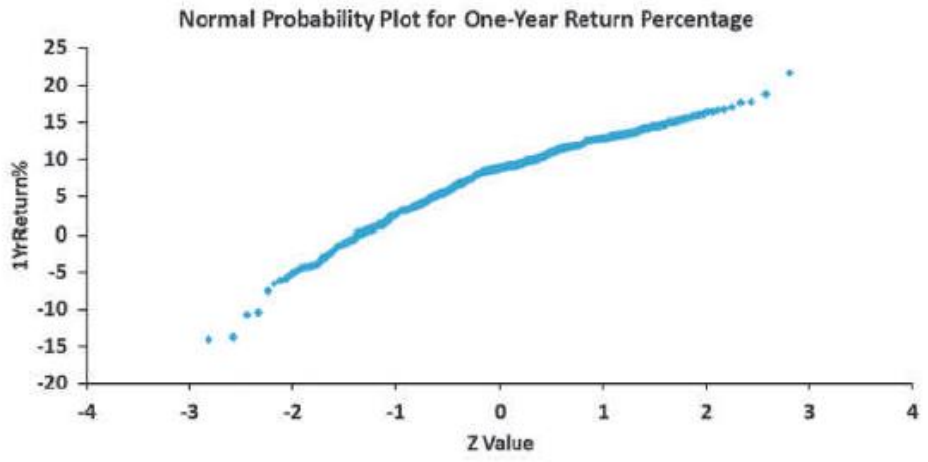
Constructing the Normal Probability Plot / Quantile–quantile plot

Step1: Arrange data into ordered array (from smallest to the largest)

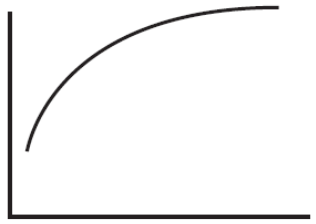
Step2: **Find corresponding standardized normal quantile values**

Step3: Plot the pairs of points with observed data values on the vertical axis and the standardized normal quantile values on the horizontal axis.

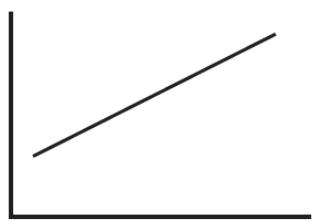
Step4: Evaluate the plot for evidence of linearity



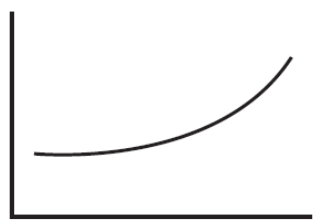
A normal probability plot for data from a normal distribution will be approximately linear:



Left-skewed
Panel A



Normal
Panel B



Right-skewed
Panel C

Run a Pilot Experiment

Setup

Homoscedasticity??
[constant variance assumption]

Objective: Compare the population variance for two **Normal distribution** independent samples

Assumptions

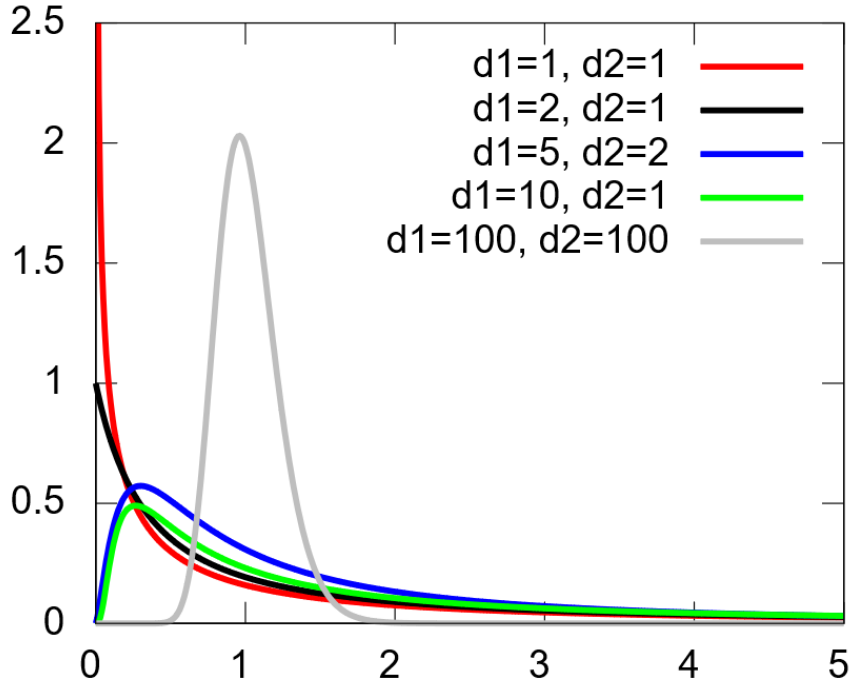
$$X_1, \cdots, X_n \sim Normal(\mu_1, \sigma_1^2)$$

$$Y_1, \cdots, Y_m \sim Normal(\mu_2, \sigma_2^2)$$

Related distribution [F-distribution]

If $X \sim \chi^2_{d_1}$ and $Y \sim \chi^2_{d_2}$ are independent, then:

$$\frac{X/d_1}{Y/d_2} \sim F(d_1, d_2)$$



$$\frac{(n-1)S_1^2}{\sigma_1^2} \sim \chi^2_{n-1}$$

$$\Rightarrow \quad \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n-1, m-1)$$

$$\frac{(m-1)S_2^2}{\sigma_2^2} \sim \chi^2_{m-1}$$

Example: Battery Experiment

Checklist: (a) Define the objectives of the experiment

Due to the frequency with which his family needed to purchase flashlight batteries, one of the authors (Dan Voss) was interested in finding out which type of nonrechargeable battery was the most economical.

In particular, Dan was interested in comparing the lifetime per unit cost of the particular name brand that he most often purchased with the store brand where he usually shopped.

He also wanted to know whether it was worthwhile paying the extra money for alkaline batteries over heavy duty batteries.

Example: Battery Experiment

Checklist: (b) Identify all sources of variation

There are several sources of variation that are easy to identify in this experiment. Clearly, different duty batteries such as alkaline and heavy duty could well be an important factor in the lifetime per unit cost, as could the brand of the battery. These two sources of variation are the ones of most interest in the experiment and form the levels of the two treatment factors "duty" and "brand."

Other possible sources of variation include the date of manufacture of the purchased battery, and whether the lifetime was monitored under continuous running conditions or under the more usual setting with the flashlight being turned on and off, the temperature of the environment, the age and variability of the flashlight bulbs.

Example: Battery Experiment

Checklist: (b) Identify all sources of variation

The first of these could not be controlled in the experiment. The batteries used in the experiment were purchased at different times and in different locations in order to give a wide representation of dates of manufacture. The variability caused by this factor would be measured as part of the natural variability (error variability) in the experiment along with measurement error. Had the dates been marked on the packets, they could have been included in the analysis of the experiment as covariates. However, the dates were not available.

The second of these possible sources of variation (running conditions) was fixed. All the measurements were to be made under constant running conditions. Although this did not mimic the usual operating conditions of flashlight batteries, Dan thought that the relative ordering of the different battery types in terms of life per unit cost would be the same. The continuous running setting was much easier to handle in an experiment since each observation was expected to take several hours and no sophisticated equipment was available.

Example: Battery Experiment

Checklist: (b) Identify all sources of variation

The third source of variation (temperature) was also fixed. Since the family living quarters are kept at a temperature of about 68◦ in the winter, Dan decided to run his experiment at this usual temperature. Small fluctuations in temperature were not expected to be important.

The variability due to the age of the flashlight bulb was more difficult to handle. A decision had to be made whether to use a new bulb for each observation and risk muddling the effect of the battery with that of the bulb, or whether to use the same bulb throughout the experiment and risk an effect of the bulb age from biasing the data. A third possibility was to divide the observations into blocks and to use a single bulb throughout a block, but to change bulbs between blocks. Since the lifetime of a bulb is considerably longer than that of a battery, Dan decided to use the same bulb throughout the experiment.

Example: Battery Experiment

Checklist: (b) Identify all sources of variation

(ii) Experimental units

The experimental units in this experiment are the time slots. These were assigned at random to the battery types so as to determine the order in which the batteries were to be observed. Any fluctuations in temperature during the experiment form part of the variability between the time slots and are included in the error variability.

(iii) Blocking factors, noise factors, and covariates

As mentioned above, it was decided not to include a blocking factor representing different flashlight bulbs. Also, the date of manufacture of each battery was not available, and small fluctuations in room temperature were not thought to be important. Consequently, there were no covariates in the experiment, and no noise factors were incorporated.

Example: Battery Experiment

Checklist: (c) Choose a rule by which to assign the experimental units to the levels of the treatment factor

Since there were to be no blocking factors, a completely randomized design was selected, and the time slots were assigned at random to the four different battery types.

Checklist: (d) Specify the measurements to be made, the experimental procedure, and the anticipated difficulties

The first difficulty was in deciding exactly how to measure lifetime of a flashlight battery. First, a flashlight requires two batteries. In order to keep the cost of the experiment low, Dan decided to wire a circuit linking just one battery to a flashlight bulb. Although this does not mimic the actual use of a flashlight, Dan thought that as with the constant running conditions, the relative lifetimes per unit cost of the four battery types would be preserved.

Example: Battery Experiment

Checklist: (d) Specify the measurements to be made, the experimental procedure, and the anticipated difficulties

Secondly, there was the difficulty in determining when the battery had run down. Each observation took several hours, and it was not possible to monitor the experiment constantly. Also, a bulb dims slowly as the battery runs down, and it is a judgment call as to when the battery is flat. Dan decided to deal with both of these problems by including a small clock in the circuit. The clock stopped before the bulb had completely dimmed, and the elapsed time on the clock was taken as a measurement of the battery life. The cost of a battery was computed as half of the cost of a two-pack, and the lifetime per unit cost was measured in minutes per dollar (min/$).

Example: Battery Experiment

Checklist: (e) Run a pilot experiment.

A few observations were run as a pilot experiment. This ensured that the circuit did indeed work properly. It was discovered that the clock and the bulb had to be wired in parallel and not in series, as Dan had first thought! The pilot experiment also gave a rough idea of the length of time each observation would take (at least four hours).

Difficulties Encountered

The only difficulty encountered in running the main experiment was that during the fourth observation, it was discovered that the clock was running but the bulb was out. This was due to a loose connection. The connection was repaired, a new battery inserted into the circuit, and the clock reset.