Stat 206: Linear Models

Lecture 15

Nov. 23, 2015

Experimental Design

- A well designed study is simple to analyze, whereas a poorly designed study often can not be rescued even with the most sophisticated analysis.
- Three principles: Ref. R.A. Fisher, The Design of Experiments, 1935.
  - *Replication*: Deals with uncertainty, allows for generalization.
  - *Randomization*: Deals with confounding factors, allows for causal-effect statements.
  - *Blocking*: Reduces known but irrelevant sources of variation, improves efficiency.

Treatment Comparison

How to compare a new drug with an old drug (status quo)?

- *Experiment A*: Apply the new drug to one patient and apply the old drug to another patient. Patient with new drug recovered, and patient with old drug did not recover. Conclusion: New drug is better.
- *Experiment B*: 200 patients were recruited. 80 patients chose to try the new drug and 120 chose to try the old drug. New drug had a 40% recovery rate and old drug had a 60% recovery rate. Conclusion: Old drug is better.
- *Experiment C*: 200 patients were recruited. 100 of them were randomly chosen and given the new drug; while the other 100 were given the old drug. New drug had a 70% recovery rate and old drug had a 50% recovery rate. Conclusion: New drug is better.

Are these conclusions well founded?

Controlled Experiments vs. Observational Studies

Data and statistical analysis could look very similar for a controlled experiment and for an observational study. However, the interpretations of the results are very different.

- In controlled experiments, treatments are randomly assigned to subjects.
  - A controlled experiment can answer whether there is a *cause-and-effect relationship* between the treatment and the response.
  - This is because randomization balances out the differences in other factors, leaving the observed differences (if any) solely attributable to the treatment.
- In observational studies, the treatment of a subject is simply observed, but not assigned by the investigator.
  - An observational study can only answer whether there is an *association* between the treatment and the response.

Experimental Design: Basic Concepts

- The goal of a study is to find out the relationships between certain explanatory factors and the response variables.
- The design of a study consists of making decisions on the following:
  - The set of response variable(s)
  - The set of explanatory factor(s)
  - The set of treatments
  - The set of experimental units
  - The way of randomization and blocking
  - Sample size and number of replications
  - The way of measuring the response variable

Factors

- Factors: Explanatory variables.
- Factor levels: “Values” of the factor.
- Types of factors:
  - Experimental factor: Levels of the factor are assigned at random to the experimental units, e.g., drug dosage.
  - Observational factor: Levels of the factor are characteristic of the experimental units and is not under control of the investigator, e.g., gender.

Treatments

- Single factor study: A treatment corresponds to a factor level. The number of treatments equals to the number of factor levels.
  - In a study of effects of education on income: each education level is a treatment (high school, college, advanced degree, etc).
- Multi-factor study: A treatment corresponds to a combination of factor levels. The number of treatments equals to the multiplication of the number of factor levels across different factors.
  - In a study of effects of race and gender on income: each combination of race and gender is a treatment (Asian female; Hispanic male, etc).

Choice of Factors and Treatments

- Which factors should be included?
- Factor levels:
  - Qualitative factors: The levels are usually indicated by the nature of the factor.
    - gender has two levels: Female and male.
  - Quantitative factors: The choice of levels reflects the type of trend expected by the investigator.
    - linear trend: two levels; quadratic trend: three levels.
    - Usually 3 to 4 equally spaced levels are sufficient.
    - The range of levels are also important.
- Often prior knowledge is required for an effective choice of factors and treatments.

Experimental Units

- An experimental unit is a subject to which a treatment is assigned.
- Experimental units should be *representative of the population* about which conclusions are going to be drawn.
  - A study conducted surveys among 5,000 US college students and found out that about 20% of them had used marijuana at least once. If the goal is to study drug usage among Americans aging from 18 to 22, is this a good design?

Sample Size and Replicates

- Sample size is the number of experimental units in the study.
  - Sample size is usually determined by the trade-off between statistical considerations such as power of tests, precision of estimations and the availability of resources such as money, time, man power, etc.
  - In general, the larger the sample size, the better for statistical inference.
- A replicate is one complete repetition of all treatments (under similar experimental conditions).
  - In a study of baking temperature on the volume of quick bread prepared from a package mix, four oven temperatures – low, medium, high and very high were tested by randomly assigning each temperature to 5 package mixes (all same brand).
  - The sample size is 20, the number of treatments is 4 and there are 5 complete replicates of the experiment.

Why Replication?

- When a treatment is repeated under the same experimental conditions, any difference observed in the responses for the same treatment is due to random errors.
- Replication provides information about the variation of the random errors.
- If this variation is relatively small compared to the total variation in the response across treatments, we would have evidence for the presence of treatment effect.
- This is the basic idea of analysis of variance.

Why Randomization?

- Randomization is introduced by R. A. Fisher in early 20th century.
- Randomization averages out between treatments whatever systematic effects may be present, apparent or hidden, so that the comparison between treatments measures only the pure treatment effect.
- Randomization is necessary not only for the assignment of treatments to experimental units, but also for other stages of the experiment, where systematic effects may be present.

What is Blocking?

- In a blocked experiment, *heterogeneous* experiment units are divided into *homogeneous* subgroups, called blocks, and separate randomization is conducted within each block.
- In a study of Vitamin C on cold prevention, 1000 children were recruited. Half of them were randomly chosen and given Vitamin C and the other half got placebos. At the end of the study, the number of colds contracted by each child was recorded. This is a *complete randomized design (CRD)*.
- If gender might also influence the incidence of cold, then a more efficient way to conduct the study is through blocking on gender:
  - 500 girls and 500 boys were recruited. Among the girls, 250 were randomly chosen and given Vitamin C and the other 250 were given placebo. Same were done for the boys. This is a *randomized complete block design (RCBD).*

Why Blocking?

- Randomization alone (as in a CRD) does not assure that the same number of girls and boys will receive each treatment. Thus if there is a difference of cold incidence rate between genders, difference between treatment groups may be observed even if there is indeed no treatment effect.
- By blocking, one removes the source of variation due to confounding factors (here gender), and thus improves the efficiency for the inference of treatment effect.

Single Factor Studies: an Example

A food company wanted to test four different package designs for a new break-fast cereal. 20 stores with approximately the same sales condition (such as sales volume, price, etc) were selected as experimental units. Five stores were randomly assigned to each of the 4 package designs.

- This is a *balanced* complete randomized design.
  - *Balanced design*: All treatments have the same number of experimental units.
  - One store was dropped from the study because of a fire: The design is not balanced anymore.
- A single, 4-level, qualitative factor: Package design.
- A quantitative response variable: Number of cereal sold during the period of study.
- Does package design affects sales? If so, how?

Package Design

Package Design (i)	Store (j)							
i	Y_{i1}	Y_{i2}	Y_{i3}	Y_{i4}	Y_{i5}	$Y_{i\cdot}$	$\overline{Y}_{i\cdot}$	n_i
1	11	17	16	14	15	73	14.6	5
2	12	10	15	19	11	67	13.4	5
3	23	20	18	17	miss	78	19.5	4
4	27	33	22	26	28	136	27.2	5
All Designs				$Y_{\cdot\cdot}$	$= 354$	$\overline{Y}_{\cdot\cdot}$	$= 18.63$	19

Single Factor ANOVA Model

This is also known as the *one-way ANOVA* model.

- Model equation:

$$Y_{ij} = \mu_i + \epsilon_{ij}, \quad i = 1, \cdots, I, \; j = 1, \cdots, n_i.$$

  - The index i denotes factor level; I is the number of factor levels (treatments).
  - The index j denotes experimental unit; n_i is the number of experimental units in the ith treatment group.
  - Y_{ij} denotes the observed outcome of the jth experimental unit in the ith treatment group.
  - μ_i denotes the ith factor level mean: These are an unknown parameters.
  - ϵ_{ij}'s denote random errors: These are unobserved.
- Model assumptions: ϵ_{ij}'s are independently and identically distributed as $Normal(0, \sigma^2)$.

Interpretation of Parameters

- Factor level mean, μ_i, stands for the mean outcome that would be obtained if the *i*th factor level were applied to the entire population from where the experimental units were sampled.
- Error variance, σ^2, stands for the variability among the outcomes if any given treatment were applied to the entire population.
- Package design.
  - The population would consist of all stores with similar sales conditions as those in the study.
  - μ_i would be the average sales of all these stores if package design *i* were used. σ^2 would be the variance of sales across all these stores when any one of the package designs were applied to all of them.

In terms of outcomes, the model says $Y_{ij} \sim_{indept} N(\mu_i, \sigma^2)$:

$$E(Y_{ij}) = \mu_i, \quad Var(Y_{ij}) = \sigma^2, \quad Cov(Y_{ij}, Y_{i'j'}) = 0, \ \ for \ \ (i,j) \neq (i',j').$$

- In words, the outcomes are independent normal random variables with a common variance and with mean equal to the respective factor level mean.
- We want to find out whether the factor level means μ_i's are all equal.
  - If $\mu_1 = \cdots = \mu_I$, then there is no treatment effect, i.e., there is no difference in different treatments in term of influencing the distribution of the outcome.
  - Otherwise, there is a treatment effect.

Estimation of Factor Level Means

By respective factor level sample means.

- $Y_{i\cdot} := \sum_{j=1}^{n_i} Y_{ij}$: Sum of outcomes in the ith treatment group.
- $\overline{Y}_{i\cdot} = \frac{1}{n_i} Y_{i\cdot}$: Sample mean of outcomes in the ith treatment group.
  - $E(\overline{Y}_{i\cdot}) = \mu_i$: $\overline{Y}_{i\cdot}$ is an unbiased estimator of μ_i.
  - $Var(\overline{Y}_{i\cdot}) = \frac{\sigma^2}{n_i}$.
  - *Notes: $\overline{Y}_{i\cdot}$ is the LS estimator (as well as MLE) of μ_i.*
- $Y_{\cdot\cdot} = \sum_{i=1}^{I} \sum_{j=1}^{n_i} Y_{ij}$: Sum of all observed outcomes.
- $\overline{Y}_{\cdot\cdot} = \frac{1}{\sum_{i=1}^{I} n_i} Y_{\cdot\cdot}$: Overall sample mean of all outcomes.

Comparison of Factor Level Means

Measure the deviation from $H_0 : \mu_1 = \mu_2 = \cdots = \mu_I$.

- A baseline value for comparison is the overall mean:

$$\mu_\cdot = \frac{1}{n_T} \sum_{i=1}^{I} n_i \mu_i, \quad n_T = \sum_{i=1}^{I} n_i.$$

- Deviation from the overall mean for each factor level:

$$(\mu_1 - \mu_\cdot)^2, \cdots, \quad (\mu_I - \mu_\cdot)^2.$$

  Under H_0, these deviations are all zero.

- Weighted sum of the deviations as an overall measurement of the deviation from H_0:

$$\sum_{i=1}^{I} n_i (\mu_i - \mu_\cdot)^2.$$

  The weight of the *ith* treatment group is its sample size n_i. This quantity is zero if and only if H_0 is true.

- Sample mean $\overline{Y}_{i\cdot}$ is an estimator of μ_i $(i = 1, \cdots, I)$.
- Overall sample mean

$$\overline{Y}_{\cdot\cdot} = \frac{1}{n_T} \sum_{i=1}^{I} \sum_{j=1}^{n_i} Y_{ij} = \frac{1}{n_T} \sum_{i=1}^{I} n_i \overline{Y}_{i\cdot}$$

  is an estimator of the overall mean $\mu_\cdot = \frac{1}{n_T} \sum_{i=1}^{I} n_i \mu_i$.
- Thus

$$\sum_{i=1}^{I} n_i (\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot})^2$$

  is an estimator of $\sum_{i=1}^{I} n_i (\mu_i - \mu_\cdot)^2$ and thus a statistic to measure the deviation from H_0.
- How to decide whether the deviation is statistically significant or not?

Decomposition of the Total Sum of Squares

Decomposition of the deviations:

$$Y_{ij} - \overline{Y}_{..} = (Y_{ij} - \overline{Y}_{i.}) + (\overline{Y}_{i.} - \overline{Y}_{..}).$$

- $Y_{ij} - \overline{Y}_{..}$: Deviation of the outcome Y_{ij} from the overall sample mean.
- $\overline{Y}_{i.} - \overline{Y}_{..}$: Deviation of the *ith* factor level sample mean from the overall sample mean.
- $Y_{ij} - \overline{Y}_{i.}$: Deviation of the outcome from its respective factor level sample mean – *residuals*.

Decomposition of the total sum of squares:

- TOtal Sum of Squares (SSTO): $\sum_{i=1}^{I} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{\cdot\cdot})^2$.
- Error Sum of Squares (SSE): $\sum_{i=1}^{I} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i\cdot})^2$.
- TReatment Sum of Squares (SSTR): $\sum_{i=1}^{I} n_i (\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot})^2$.

Decomposition of the total sum of squares:

$$
\begin{aligned}
SSTO &= \sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{\cdot\cdot})^2 \\
&= \sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i\cdot})^2 + \sum_{i=1}^{I} n_i(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot})^2 \\
&= SSE + SSTR.
\end{aligned}
$$

- TOtal Sum of Squares (SSTO): $\sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{\cdot\cdot})^2$.
- Error Sum of Squares (SSE): $\sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i\cdot})^2$.
- TReatment Sum of Squares (SSTR): $\sum_{i=1}^{I} n_i(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot})^2$.

$$df(SSTO) = df(SSTR) + df(SSE).$$

- The degrees of freedom of SSTO is , since

$$\sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{\cdot\cdot}) = 0.$$

- The degrees of freedom of SSTR is , since

$$\sum_{i=1}^{I} n_i(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot}) = 0.$$

- The degrees of freedom of SSE is
, since for each $i = 1, \cdots, I$

$$\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i\cdot}) = 0.$$

$$df(SSTO) = df(SSTR) + df(SSE).$$

- The degrees of freedom of SSTO is $n_T - 1$, since

$$\sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{..}) = 0.$$

- The degrees of freedom of SSTR is $I - 1$, since

$$\sum_{i=1}^{I} n_i(\overline{Y}_{i.} - \overline{Y}_{..}) = 0.$$

- The degrees of freedom of SSE is $\sum_{i=1}^{I}(n_i - 1) = n_T - I$, since for each $i = 1, \cdots, I$

$$\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i.}) = 0.$$

Interpretation of SS

- SSTO: A measure of the overall variability among the outcomes.
- SSTR: A measure of the variability among the . The more similar the factor level means, the SSTR tends to be.
- SSE: A measure of the variation of the outcomes around their respective factor level means. The smaller the , the smaller SSE tends to be.

Overall variability of the outcome is the sum of the variability due to different treatments and the variability due to random fluctuations.

Interpretation of SS

- SSTO: A measure of the overall variability among the outcomes.
- SSTR: A measure of the variability among the factor level means. The more similar the factor level means, the smaller SSTR tends to be.
- SSE: A measure of the variation of the outcomes around their respective factor level means. The smaller the error variance, the smaller SSE tends to be.

Overall variability of the outcome is the sum of the variability due to different treatments and the variability due to random fluctuations.

Package Design

- $SSTO =$
  $(11 - 18.63)^2 + (17 - 18.63)^2 + \cdots + (28 - 18.63)^2 = 746.42$
- $SSTR = 5(14.6 - 18.63)^2 + 5(13.4 - 18.63)^2 + 4(19.5 - 18.63)^2 + 5(27.2 - 18.63)^2 = 588.22$
- $SSE = \{(11 - 14.6)^2 + \cdots + (15 - 14.6)^2\} + \cdots + \{(27 - 27.2)^2 + \cdots + (28 - 27.2)^2\} = 158.20$
- Check: $746.42 = 588.22 + 158.20$

Sampling Distributions of SS

- For each $i = 1, \cdots, I$, $\{Y_{ij}\}_{j=1}^{n_i}$ are i.i.d. $N(\mu_i, \sigma^2)$, thus

$$\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i\cdot})^2 \sim \sigma^2 \chi^2_{(n_i-1)}.$$

- Since $\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i\cdot})^2$ are independent across i, so

$$SSE = \sum_{i=1}^{I}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i\cdot})^2 \sim \sigma^2 \chi^2_{(n_T-I)}.$$

  Note, $n_T - I = \sum_{i=1}^{I}(n_i - 1)$ is the degrees of freedom of SSE.

- Therefore,

$$E[SSE] = (n_T - I)\sigma^2.$$

SSTR is independent with *SSE* (*Why?*) and $E[SSTR] = (I-1)\sigma^2 + \sum_{i=1}^{I} n(\mu_i - \mu_\cdot)^2$.

- For simplicity, let's consider a balanced design, i.e., $n_1 = \cdots = n_I = n$.
- $\overline{Y}_{i\cdot} - \mu_i$ are i.i.d. $N(0, \sigma^2/n)$ and $\frac{1}{I}\sum_{i=1}^{I}(\overline{Y}_{i\cdot} - \mu_i) = \bar{Y}_{\cdot\cdot} - \mu_\cdot$.
- Thus $\sum_{i=1}^{I}\left\{(\bar{Y}_{i\cdot} - \mu_i) - (\bar{Y}_{\cdot\cdot} - \mu_\cdot)\right\}^2 \sim \frac{\sigma^2}{n}\chi^2_{(I-1)}$ and

$$E\left(\sum_{i=1}^{I}\left\{(\bar{Y}_{i\cdot} - \mu_i) - (\bar{Y}_{\cdot\cdot} - \mu_\cdot)\right\}^2\right) = (I-1)\frac{\sigma^2}{n}$$

$$= \sum_{i=1}^{n}(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2 - \sum_{i=1}^{n}(\mu_i - \mu_\cdot)^2.$$

Moreover, under H_0 : $SSTR \sim \sigma^2\chi^2_{(I-1)}$.

Expectation of sum of squares:

- $E(SSE) = (n_T - I)\sigma^2$.
- $E(SSTR) = (I-1)\sigma^2 + \sum_{i=1}^{I} n_i(\mu_i - \mu_\cdot)^2$.
- $E(SSTO) = (n_T - 1)\sigma^2 + \sum_{i=1}^{I} n_i(\mu_i - \mu_\cdot)^2$.

Under $H_0 : \mu_1 = \mu_2 = \cdots = \mu_I$:

- $E(SSE) = (n_T - I)\sigma^2$, and $SSE \sim \sigma^2\chi^2_{(n_T-I)}$.
- $E(SSTR) = (I-1)\sigma^2$, and $SSTR \sim \sigma^2\chi^2_{(I-1)}$.
- $E(SSTO) = (n_T - 1)\sigma^2$, and $SSTO \sim \sigma^2\chi^2_{(n_T-1)}$.

Also, SSTR and SSE are independent (whether under H_0 or not).

F-test for Equality of Means

- Mean squares: MS: = SS/df.

$$MSTR := \frac{SSTR}{I-1}, \quad MSE := \frac{SSE}{n_T - I}.$$

- F-ratio:

$$F^* = \frac{\frac{SSTR}{I-1}}{\frac{SSE}{n_T-I}} = \frac{MSTR}{MSE}$$

- Under H_0: $F^* \sim_{H_0} F_{(I-1),(n_T-I)}$.
- At significance level α:
  - P-value approach: Calculate the p-value $p = P(F_{I-1,n_T-I} > F^*)$. If $p < \alpha$, reject H_0.
  - Critical value approach: Find the critical value $F(1-\alpha; I-1, n_T - I)$. If $F^* > F(1-\alpha; I-1, n_T - I)$, reject H_0.

Anova Table for Single Factor Models

Source of Variation	Sum of Squares (SS)	Degrees of Freedom (df)	MS $= \frac{SS}{df}$	E(MS)
Between treatments	SSTR	$I - 1$	MSTR	$\sigma^2 + \frac{\sum_{i=1}^{I} n_i(\mu_i - \mu_.)^2}{I-1}$
Within treatments	SSE	$n_T - I$	MSE	σ^2
Total	SSTO	$n_T - 1$		

Package Design

Source of Variation	Sum of Squares (SS)	Degrees of Freedom (df)	MS
Between treatments	SSTR=588.22	$I - 1 = 3$	MSTR=196.07
Within treatments	SSE=158.20	$n_T - I = 15$	MSE=10.55
Total	SSTO=746.42	$n_T - 1 = 18$	

- F-ratio

$$F^* = \frac{MSTR}{MSE} = \frac{196.07}{10.55} = 18.59.$$

- p-value

$$P(F_{3,15} > 18.59) < 0.0001.$$

- The result is highly significant, so we should reject $H_0 : \mu_1 = \cdots = \mu_4$.

Analysis of Factor Level Means

Upon the rejection of $H_0 : \mu_1 = \cdots = \mu_I$, we want to investigate the nature of the differences among the factor level means:

- Factor level means: μ_i.
- Differences between two factor level means: $D = \mu_i - \mu_j$.
- *Contrasts* of factor level means: $L = \sum_{i=1}^{I} c_i \mu_i$, where $\sum_{i=1}^{I} c_i = 0$.

When more than one mean/contrast are involved, we also need to consider procedures that account for *multiple comparisons*:

- Bonferroni's procedure
- Tukey's procedure
- Scheffe's procedure

Factor Level Mean

Factor level sample mean $\overline{Y}_{i\cdot}$ as a point estimator of μ_i:

- It is an estimator: $E(\overline{Y}_{i\cdot}) =$.
- $\mathrm{Var}(\overline{Y}_{i\cdot}) =$.
- Standard error of $\overline{Y}_{i\cdot}$:

$$s(\overline{Y}_{i\cdot}) = \sqrt{\frac{MSE}{n_i}}.$$

- Pivotal quantity:

$$\frac{\overline{Y}_{i\cdot} - \mu_i}{\sqrt{\frac{MSE}{n_i}}} \sim .$$

- $(1 - \alpha)$-confidence interval of μ_i:

$$\overline{Y}_{i\cdot} \pm s(\overline{Y}_{i\cdot}) t(1 - \frac{\alpha}{2}; n_T - I).$$

Factor Level Mean

Factor level sample mean $\overline{Y}_{i\cdot}$ as a point estimator of μ_i:

- It is an unbiased estimator: $E(\overline{Y}_{i\cdot}) = \mu_i$.
- $\mathrm{Var}(\overline{Y}_{i\cdot}) = \frac{\sigma^2}{n_i}$.
- Standard error of $\overline{Y}_{i\cdot}$:

$$s(\overline{Y}_{i\cdot}) = \sqrt{\frac{MSE}{n_i}}.$$

- Pivotal quantity:

$$\frac{\overline{Y}_{i\cdot} - \mu_i}{\sqrt{\frac{MSE}{n_i}}} \sim t_{(n_T - I)}.$$

- $(1 - \alpha)$-confidence interval of μ_i:

$$\overline{Y}_{i\cdot} \pm s(\overline{Y}_{i\cdot}) t(1 - \frac{\alpha}{2}; n_T - I).$$

Test $H_0 : \mu_i = c$ vs. $H_a : \mu_i \neq c$.

- T-statistic:

$$T^* = \frac{\overline{Y}_{i\cdot} - c}{\sqrt{\frac{MSE}{n_i}}}.$$

- Under H_0: $T^* \sim t_{(n_T - I)}$.
- At significance level α, reject H_0 if $|T^*| > t(1 - \frac{\alpha}{2}; n_T - I)$.
- *Confidence interval approach*: If c does not belong to the $(1 - \alpha)$ - C.I of μ_i, then reject H_0 at level α.
  Notes: Confidence interval approach is equivalent to Pvalue approach and critical value approach. Why?

Package Design

Package Design (i)	Store (j)							
i	Y_{i1}	Y_{i2}	Y_{i3}	Y_{i4}	Y_{i5}	$Y_{i\cdot}$	$\overline{Y}_{i\cdot}$	n_i
1	11	17	16	14	15	73	14.6	5
2	12	10	15	19	11	67	13.4	5
3	23	20	18	17	miss	78	19.5	4
4	27	33	22	26	28	136	27.2	5
All Designs				$Y_{\cdot\cdot}$	$= 354$	$\overline{Y}_{\cdot\cdot}$	$= 18.63$	19

- $\overline{Y}_{1\cdot} = 14.6$, $MSE = 10.55$, $df(MSE) = 19 - 4 = 15$, $n_1 = 5$.
- $s(\overline{Y}_{1\cdot}) = \sqrt{MSE/n_1} = \sqrt{10.55/5} = 1.45$.
- A 95% confidence interval of μ_1:

$$14.6 \pm 1.45 \times t(1 - \frac{0.05}{2}; 15) = 14.6 \pm 1.45 \times 2.131$$
$$= [11.51, 17.69].$$

Difference Between Two Means

$D = \mu_i - \mu_j$ for some $i \neq j$.

- $\widehat{D} =$ is an estimator of D.
- $\mathrm{Var}(\widehat{D}) =$
- $\frac{\widehat{D}-D}{s(\widehat{D})} \sim t_{(n_T-I)}$. $(1-\alpha)$ - confidence interval of D:

$$\widehat{D} \pm s(\widehat{D}) t(1 - \frac{\alpha}{2}; n_T - I).$$

- Test $H_0 : D = 0$ vs. $H_a : D \neq 0$. At the significance level α, check whether

$$0 \in \widehat{D} \pm s(\widehat{D}) t(1 - \frac{\alpha}{2}; n_T - I).$$

If , reject H_0 at level α and conclude the two means are different.

Difference Between Two Means

$D = \mu_i - \mu_j$ for some $i \neq j$.

- $\widehat{D} = \overline{Y}_{i\cdot} - \overline{Y}_{j\cdot}$ is an unbiased estimator of D.
- $\mathrm{Var}(\widehat{D}) = \mathrm{Var}(\overline{Y}_{i\cdot}) + \mathrm{Var}(\overline{Y}_{j\cdot}) = \sigma^2\{\frac{1}{n_i} + \frac{1}{n_j}\}$
- $s(\widehat{D}) = \sqrt{MSE(1/n_i + 1/n_j)}$.
- $\frac{\widehat{D}-D}{s(\widehat{D})} \sim t_{(n_T-I)}$. $(1-\alpha)$ - confidence interval of D:

$$\widehat{D} \pm s(\widehat{D})t(1-\frac{\alpha}{2}; n_T - I).$$

- Test $H_0 : D = 0$ vs. $H_a : D \neq 0$. At the significance level α, check whether

$$0 \in \widehat{D} \pm s(\widehat{D})t(1-\frac{\alpha}{2}; n_T - I).$$

If not, reject H_0 at level α and conclude the two means are different.

Rust Inhibitors

In a study of the effectiveness of different rust inhibitors, four brands (1,2,3,4) were tested. Altogether, 40 experimental units were randomly assigned to the four brands, with 10 units assigned to each brand. The resistance to rust was evaluated in a coded form after exposing the experimental units to severe conditions. This is a *balanced complete randomized design (CRD)*.
Summary statistics and ANOVA table: $n_1 = n_2 = n_3 = n_4 = 10$ and $\overline{Y}_{1\cdot} = 43.14$, $\overline{Y}_{2\cdot} = 89.44$, $\overline{Y}_{3\cdot} = 67.95$, $\overline{Y}_{4\cdot} = 40.47$.

Source of Variation	Sum of Squares (SS)	Degrees of Freedom (df)	MS
Between treatments	SSTR=15953.47	$I - 1 = 3$	MSTR=5317.82
Within treatments	SSE=221.03	$n_T - I = 36$	MSE=6.140
Total	SSTO=16174.50	$n_T - 1 = 39$	

95% C.I and testing for $D = \mu_1 - \mu_2$.

- $\widehat{D} = 43.14 - 89.44 = -46.3$.
- $s(\widehat{D}) = \sqrt{MSE(\frac{1}{n_1} + \frac{1}{n_2})} = \sqrt{6.14 \times \frac{2}{10}} = 1.11$.
- $t(1 - \frac{\alpha}{2}; n_T - I) = t(0.975; 36) = 2.03$.
- 95% C.I: $-46.3 \pm 1.11 \times 2.03 = [-48.6, \ -44]$.
- Since $0 \notin [-48.6, \ -44]$, reject $H_0 : \mu_1 = \mu_2$ at the 0.05 significance level.