Stat 206: Linear Models

Lecture 17

Nov. 27, 2019

Experimental Design: Basic Principles

Three basic principles: *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.

Experimental Design: Basic Concepts

- Goal: Find out the relationships between explanatory factors and response variables and make cause-effect statements (if applicable).
- Experimental design 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 are assigned at random to the experimental units, e.g., drug dosage.
  - Observational factor: Levels are characteristic of the experimental units and is not under control of the investigator, e.g., gender.

Treatments

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: Levels are indicated by the nature of the factor.
    - gender: Female and male.
  - Quantitative factors: Levels reflect the expected trend.
    - 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
about which conclusions are going to be drawn.

Sample Size and Replicates

- Sample size is the number of experimental units in the study.
  - 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.

Single Factor Study: 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 (CRD).
  - *Balanced design*: All treatments have the same number of experimental units.
  - Later 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

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

- 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. $n_T = \sum_{i=1}^{I} n_i$ is the total size sample.
- 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 r.v.s and assumed to be i.i.d. $Normal(0, \sigma^2)$.

- The outcomes are random variables with variance and with means equal to:

$$E(Y_{ij}) = \quad , \quad Var(Y_{ij}) = \quad , \quad Cov(Y_{ij}, Y_{i'j'}) = \quad , \ for \ (i,j) \neq (i',j')$$

- We want to find out whether the factor level means μ_i's are
  - If $\mu_1 = \cdots = \mu_I$, then there is treatment effect, i.e., the outcomes have the same distribution under different treatments.
  - Otherwise, there is a treatment effect.

Decomposition of the Total Deviations

Decomposition of the deviations:

- $Y_{ij} - \overline{Y}_{..}$: 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* e_{ij}.

Decomposition of the Total Sum of Squares

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

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

- 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 $i = 1, \cdots, I$

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

Interpretation of SS

- SSTO: A measure of the 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 equals to the sum of the variability due to different treatments and the variability due to random fluctuations.

Sampling Distributions of SS

- $SSE = \sum_{i=1}^{I} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i\cdot})^2 \sim \sigma^2 \chi^2_{(n_T - I)}$.
- $SSTR$ is independent with SSE
- $E[SSTR] = (I-1)\sigma^2 + \sum_{i=1}^{I} n_i(\mu_i - \mu_\cdot)^2$ where $\mu_\cdot = \frac{1}{n_T}\sum_{i=1}^{I} n_i \mu_i$.
- Under $H_0 : \mu_1 = \cdots = \mu_I$, $SSTR \sim \sigma^2 \chi^2_{(I-1)}$.
- Mean squares: MS: = SS/df.

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

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	$\sigma^2 + \frac{\sum_{i=1}^{I} n_i(\mu_i-\mu_\cdot)^2}{I-1}$
Within treatments	SSE	$n_T - I$	MSE	σ^2
Total	SSTO	$n_T - 1$		

F-test for Equality of Means

- 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.

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

- Investigate the nature of the differences among the factor level means.
- Comparison between two factor level means: $D = \mu_i - \mu_j$.
- *Contrast* of factor level means: $L = \sum_{i=1}^{I} c_i \mu_i$, where $\sum_{i=1}^{I} c_i = 0$.

Comparison Between Two Means

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

- $\widehat{D} =$ is an estimator of D.
- $\mathrm{Var}(\widehat{D}) =$, $s(\widehat{D}) =$
- $\frac{\widehat{D}-D}{s(\widehat{D})} \sim$: $(1-\alpha)$ - 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.

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)*.

$\overline{Y}_{1\cdot} = 43.14,\ \overline{Y}_{2\cdot} = 89.44,\ \overline{Y}_{3\cdot} = 67.95,\ \overline{Y}_{4\cdot} = 40.47.$

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.

Multiple Comparison

A family of statistical inferences are considered **simultaneously**:

- Errors are more likely to occur.
  - Suppose one tests 100 null hypotheses which are indeed all true. If the type I error rate of each test is 5% and if these tests are independent, then the probability of making at least one false rejection is $1 - 0.95^{100} = 99.4\%$.
- Simultaneously control the probability of committing such errors.
  - Multiple hypothesis testing: Control the family-wise type-I error rate (FWER).
  - Simultaneous confidence region: Maintain a family-wise confidence level.

Family-wise Confidence Intervals for Pairwise Comparisons

- For I factor levels, there are $I(I-1)/2$ distinct pairwise comparisons of the form $D_{ij} = \mu_i - \mu_j$ $(1 \leq i < j \leq I)$.
- Denote the $(1-\alpha)$-C.I. for D_{ij} by $C_{ij}(\alpha)$:

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

- $t(1 - \frac{\alpha}{2}; n_T - I)$ is the multiplier that gives the desired confidence coefficient $1 - \alpha$:

$$P(D_{ij} \in C_{ij}(\alpha)) = 1 - \alpha,$$

i.e., the probability that D_{ij} falls out of C_{ij} is at most α.

- *Family-wise confidence coefficient* of this family of confidence intervals is defined as:

  i.e., the probability that these C.Is cover their respective parameter.

- $$P(D_{ij} \in C_{ij}(\alpha), \text{ for all } 1 \leq i < j \leq I)$$
  $$P(D_{ij} \in C_{ij}(\alpha)) = 1 - \alpha.$$

- How to construct C.Is such that the family-wise confidence coefficient is at least $1 - \alpha$?
- We should replace $t(1 - \frac{\alpha}{2}; n_T - I)$ by a multiplier (resulting in C.Is).