10.2 An Overview

the sample the variable is from

One-Way Analysis of Variance : the mean of a random variable depends only on a single factor

Two-Way Analysis of Variance : the mean of a random variable depends on two factors

- The analysis of variance approach for testing a null hypothesis H_0 concerning multiple parameters relating to the population means is based on deriving two estimators of the common variance σ^2.

10.3 One-Way Analysis of Variance

- we will be testing the null hypothesis that all the population means (for m treatments) are equal against the alternative that at least two of them differ

- Since there are a total of nm independent normal random variables X_{ij}, it follows that the sum of squares of their standardized versions will be a χ^2 random variable with nm degrees of freedom.

$$\sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij}-E[X_{ij}])^2/\sigma^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij}-\mu_i)^2/\sigma^2 \sim \chi^2_{nm}$$

$$\Rightarrow \sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij}-E[X_{ij}])^2/\sigma^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij}-X_{i.})^2/\sigma^2 \sim \chi^2_{nm-m}$$

First Estimator of σ^2 (No assumptions needed)

- the expected value of a χ^2 random variable is equal to its degrees of freedom, that is to say,

$$E\Big[\underbrace{\sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij}-E[X_{ij}])^2}_{SS_W}/\sigma^2\Big] = E[\chi^2_{nm-m}] = nm-m \iff E[SS_W/(nm-m)] = \sigma^2$$

"Within Samples Sum of Squares"

☆ note that this is in conclusion, our first estimator of σ^2, and it was obtained without any assumption

Second Estimator of σ^2 (under the assumption H_0 is true)

- Given H_0 is true, m sample means will all be normally distributed with the same mean μ and the same variance σ^2/n. Hence, the sum of squares of the m standardized variables will be χ^2 random variable with m d.f.

$$\frac{X_{i.}-\mu}{\sqrt{\sigma^2/n}} = \sqrt{n}(X_{i.}-\mu)/\sigma \Rightarrow n\sum_{i=1}^{m}(X_{i.}-\mu)^2/\sigma^2 \sim \chi^2_m$$

$$\Rightarrow \frac{X_{i.}-X_{..}}{\sqrt{\sigma^2/n}} = \sqrt{n}(X_{i.}-X_{..})/\sigma \Rightarrow \underbrace{n\sum_{i=1}^{m}(X_{i.}-X_{..})^2}_{SS_b}/\sigma^2 \sim \chi^2_{m-1}$$, this results in,

"Between Samples Sum of Squares"

$$E[SS_b]/\sigma^2 = m-1$$

$$\Rightarrow E[SS_b/m-1] = \sigma^2$$

Thus we have shown that

$SS_W/(nm-m)$	always estimates σ^2
$SS_b/(m-1)$	estimates σ^2 when H_0 is true

Because it can be shown that $SS_b/(m-1)$ will tend to exceed σ^2 when H_0 is not true,* it is reasonable to let the test statistic be given by

$$TS = \frac{SS_b/(m-1)}{SS_W/(nm-m)}$$

and to reject H_0 when TS is sufficiently large.

☆ When H_0 is true, TS has an F-distribution with $m-1$ and $m(n-1)$ degrees of freedom.

TABLE 10.2 *One-Way ANOVA Table*

Source of Variation	Sum of Squares	Degrees of Freedom	Value of Test Statistic
Between samples	$SS_b = n\sum_{i=1}^{m}(X_{i.} - X_{..})^2$	$m-1$	
Within samples	$SS_W = \sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij} - X_{i.})^2$	$nm-m$	$TS = \frac{SS_b/(m-1)}{SS_W/(nm-m)}$

Significance level α test:
reject H_0 if $TS \geq F_{m-1,nm-m,\alpha}$
do not reject otherwise

If $TS = v$, then p-value $= P\{F_{m-1,nm-m} \geq v\}$

10.3.1 Multiple Comparisons of Sample Means

10.3.2 One-Way Analysis of Variance with Unequal Sample Sizes

$$\sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij}-X_{i.})^2/\sigma^2 \sim \chi^2_{\sum_{i=1}^{m} n_i - m}$$

$$\Rightarrow \underbrace{\sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij}-X_{i.})^2}_{SS_W}\Big/\left(\sum_{i=1}^{m} n_i - m\right) \text{ is an unbiased estimator of } \sigma^2$$

$$\underbrace{\sum_{i=1}^{m} n_i(X_{i.}-X_{..})^2}_{SS_b}/\sigma^2 = m-1 \Rightarrow \sum_{i=1}^{m} n_i(X_{i.}-X_{..})^2/(m-1) = \sigma^2 \text{ is the second unbiased estimator of } \sigma^2$$

Consequently,

$$\frac{SS_b/(m-1)}{SS_W/(\sum_{i=1}^{m} n_i - m)}$$ is an F-random variable with $m-1$ and $\sum_{i=1}^{m} n_i - m$ degrees of freedom.

REMARK

When the samples are of different sizes we say that we are in the *unbalanced* case. Whenever possible it is advantageous to choose a balanced design over an unbalanced one. For one thing, the test statistic in a balanced design is relatively insensitive to slight departures from the assumption of equal population variances. (That is, the balanced design is more robust than the unbalanced one.)

10.4 Two-Factor Analysis of Variance : Introduction and Parameter Estimation

$\mu_{ij} = \mu + \alpha_i + \beta_j$, where $\mu = E(X_{..})$, $E(X_{i.} - X_{..}) = \alpha_i$, $E(X_{.j} - X_{..}) = \beta_j$

10.5 Two-Factor Analysis of Variance : Testing Hypothesis

To obtain our first estimator of σ^2, we start with the fact that

$$\sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij} - E[X_{ij}])^2/\sigma^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij} - \mu - \alpha_i - \beta_j)^2/\sigma^2$$

is χ^2 with nm degrees of freedom. We now replace the unknown parameters by their estimators $\hat{\mu}$, $\hat{\alpha}_1$, $\hat{\alpha}_2, \ldots, \hat{\alpha}_m$, $\hat{\beta}_1, \ldots, \hat{\beta}_n$, then it turns out that the resulting expression will remain χ^2 but will lose 1 degree of freedom for each parameter that is estimated.

↑ read pdf 526-527 for detailed explanation about degrees of freedom

First Estimator of σ^2

Since $\hat{\mu} = X_{..}, \hat{\alpha}_i = X_{i.} - X_{..}, \hat{\beta}_j = X_{.j} - X_{..}$, it follows that $\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j = X_{i.} + X_{.j} - X_{..}$; thus,

$$\sum_{i=1}^{m} \sum_{j=1}^{n} (X_{ij} - X_{i.} - X_{.j} + X_{..})^2/\sigma^2 \tag{10.5.1}$$

is a chi-square random variable with $(n-1)(m-1)$ degrees of freedom.

Definition

The statistic SS_e defined by

$$SS_e = \sum_{i=1}^{m} \sum_{j=1}^{n} (X_{ij} - X_{i.} - X_{.j} + X_{..})^2$$

is called the *error sum of squares*.

$\Rightarrow E[SS_e/\sigma^2] = (n-1)(m-1)$

$\approx E[SS_e/(n-1)(m-1)] = \sigma^2$ is an unbiased estimator of σ^2

Second Estimator of σ^2

- Note that if H_0 is true, $E(X_{i.}) = \mu + \alpha_i = \mu$, and because each $X_{i.}$ is the average of n random variables, each having variance σ^2, it follows $Var(X_{i.}) = \frac{\sigma^2}{n}$

$\Rightarrow \sum_{i=1}^{M} [X_{i.} - E(X_{i.})^2] / Var(X_{i.}) = n \sum_{i=1}^{m} (X_{i.} - \mu)^2/\sigma^2$ will be χ^2 with m degrees of freedom.

$\approx \underbrace{n \sum_{i=1}^{m} (X_{i.} - X_{..})^2}_{SS_r}/\sigma^2$

$\Rightarrow E[SS_r/\sigma^2] = m-1$

$E[SS_r/m-1] = \sigma^2$ is an unbiased estimator.

☆ $SS_r/m-1$ will tend to be larger than σ^2 when H_0 is not true.

TABLE 10.3 *Two-Factor ANOVA*

	Sum of Squares	Degrees of Freedom
Row	$SS_r = n\sum_{i=1}^{m}(X_{i.} - X_{..})^2$	$m-1$
Column	$SS_c = m\sum_{j=1}^{n}(X_{.j} - X_{..})^2$	$n-1$
Error	$SS_e = \sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij} - X_{i.} - X_{.j} + X_{..})^2$	$(n-1)(m-1)$

Let $N = (n-1)(m-1)$

Null Hypothesis	Test Statistic	Significance Level α Test	p-value if $TS = v$
All $\alpha_i = 0$	$\frac{SS_r/(m-1)}{SS_e/N}$	Reject if $TS \geq F_{m-1,N,\alpha}$	$P\{F_{m-1,N} \geq v\}$
All $\beta_j = 0$	$\frac{SS_c/(n-1)}{SS_e/N}$	Reject if $TS \geq F_{n-1,N,\alpha}$	$P\{F_{n-1,N} \geq v\}$

10.6 Two-Way Analysis of Variance With Interaction

$\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$, where $\mu = \mu_{..}$, $\alpha_i = \mu_{i.} - \mu_{..}$, $\beta_j = \mu_{.j} - \mu_{..}$, $\gamma_{ij} = \mu_{ij} - \mu_{i.} - \mu_{.j} - \mu$

* γ_{ij} is the amount by which μ_{ij} exceeds the sum of the grand mean and the increments due to row i and to column j.
It is a measure of departure from row and column additivity of the mean value μ_{ij}, and is called the interaction of row i and column j

$E[X_{ijk}] = \mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$

$E[X_{ij.}] = \mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$

$E[X_{i..}] = \mu + \alpha_i$

$E[X_{.j.}] = \mu + \beta_j$

$E[X_{...}] = \mu$, therefore, the unbiased estimators are given by

$\hat{\mu} = X_{...}$

$\hat{\alpha}_i = X_{i..} - X_{...}$

$\hat{\beta}_j = X_{.j.} - X_{...}$

$\hat{\gamma}_{ij} = X_{ij.} - \hat{\mu} - \hat{\beta}_j - \hat{\alpha}_i = X_{ij.} - X_{i..} - X_{.j.} + X_{...}$

$\approx \sum_{k=1}^{l} \sum_{j=1}^{n} \sum_{i=1}^{m} \frac{\overbrace{(X_{ijk} - \mu - \alpha_i - \beta_j - \gamma_{ij})}^{\text{residual}}{}^2}{\sigma^2}$ is a χ^2 random variable with nml degrees of freedom

$\Rightarrow$ let $SS_e = \sum_{k=1}^{l} \sum_{j=1}^{n} \sum_{i=1}^{m} (X_{ijk} - X_{ij.})^2$, SS_e/σ^2 is χ^2 with $nm(l-1)$ degrees of freedom,

$\Rightarrow \frac{SS_e}{nm(l-1)}$ is an unbiased estimator of σ^2

- Under the condition $H_0: \gamma_{ij} = 0$ is true, $E(X_{ij.}) = \mu + \alpha_i + \beta_j \Rightarrow Var(X_{ij.}) = \sigma^2/l$.
Hence, under the assumption of no interaction, $\sum_{j=1}^{n} \sum_{i=1}^{m} \frac{l(X_{ij.} - \mu - \alpha_i - \beta_j)^2}{\sigma^2}$ is a χ^2 random variable with nm degrees of freedom.

$\Rightarrow \underbrace{\sum_{j=1}^{n} \sum_{i=1}^{m} l(X_{ij.} - X_{i..} - X_{.j.} + X_{...})^2}_{SS_{int}} / \sigma^2$ is χ^2 with $(n-1)(m-1)$ degrees of freedom.

$\Rightarrow SS_{int}/(n-1)(m-1)$ is an unbiased estimator of σ^2

TABLE 10.4 *Two-way ANOVA with l Observations per Cell: $N = nm(l-1)$*

Source of Variation	Degrees of Freedom	Sum of Squares	F-Statistic	Level α Test	p-Value if $F = v$
Row	$m-1$	$SS_r = ln\sum_{i=1}^{m}(X_{i..} - X_{...})^2$	$F_r = \frac{SS_r/(m-1)}{SS_e/N}$	Reject H_0^r if $F_r > F_{m-1,N,\alpha}$	$P\{F_{m-1,N} > v\}$
Column	$n-1$	$SS_c = lm\sum_{j=1}^{n}(X_{.j.} - X_{...})^2$	$F_c = \frac{SS_c/(n-1)}{SS_e/N}$	Reject H_0^c if $F_c > F_{n-1,N,\alpha}$	$P\{F_{n-1,N} > v\}$
Interaction	$(n-1)(m-1)$	$SS_{int} = l\sum_{j=1}^{n} \times \sum_{i=1}^{m}(X_{ij.} - X_{i..} - X_{.j.} + X_{...})^2$	$F_{int} = \frac{SS_{int}/(n-1)(m-1)}{SS_e/N}$	Reject H_0^{int} if $F_{int} > F_{(n-1)(m-1),N,\alpha}$	$P\{F_{(n-1)(m-1),N} > v\}$
Error	N	$SS_e = \sum_{k=1}^{l}\sum_{j=1}^{n} \times \sum_{i=1}^{m}(X_{ijk} - X_{ij.})^2$			