Overview

- This is a summary of what we have learned in this semester. The following slides do not cover everything.
- Materials mentioned in this tutorial will just assist you to prepare your own summary for the final exam.

One-way ANOVA Model

We denote sampled data values as Y_{ij}, where $i = 1, \ldots, k$ indicates the factor level and $j = 1, \ldots, n_i$ indicates a specific value within the i^{th} factor level. We might write:

$$Y_{ij} = \mu + \tau_i + \varepsilon_i,$$

with some constraints to avoid overparameterisation. Here τ_i is the i^{th} level effect or treatment effect.

- Treatment contrasts. $\tau_1 = 0$
- Sum contrasts. $\sum_{i=1}^{k} n_i \tau_i = 0$

The two parameterisations have different formats of estimators of μ_i and τ_i (Page 3-5 of Lecture Brick).

Contrast of μ_i's

We can find a 100(1-α)% confidence interval for any linear combination of the μ_i's, say $h_1\mu_1 + \cdots + h_k\mu_k$, for any vector of constants $h = (h_1, \ldots, h_k)$. Such a linear combination is often called a contrast.

Since normally "within factor" averages are formed from disjoint (and therefore independent) subsets of the observed responses, we have $\bar{Y}_i$'s are independent. Then we have

$$Var(\sum_{i=1}^{k} h_i \bar{Y}_i) = \sum_{i=1}^{k} h_i^2 Var(\bar{Y}_i) = \sigma^2 \sum_{i=1}^{k} \frac{h_i^2}{n_i}.$$

Contrast of μ_i's

Thus, the desired confidence interval would be

$$(\sum_{i=1}^{k} h_i \bar{Y}_i) \pm t_{n-k}(1 - \frac{\alpha}{2}) s \sqrt{\sum_{i=1}^{k} \frac{h_i^2}{n_i}}.$$

We can also test hypotheses of the form:

$$H_0 : \sum_{i=1}^{k} h_i \mu_i = c_0 \quad versus \quad H_0 : \sum_{i=1}^{k} h_i \mu_i \neq c_0.$$

Using the test statistic:

$$T = \frac{\sum_{i=1}^{k} h_i \bar{Y}_i - c_0}{s \sqrt{\sum_{i=1}^{k} \frac{h_i^2}{n_i}}}.$$

Random Effects

A one-way ANOVA model with random effects

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

where $\alpha_i \overset{i.i.d.}{\sim} \mathrm{Normal}(0, \sigma_\alpha^2)$ and $\varepsilon_{ij} \overset{i.i.d.}{\sim} \mathrm{Normal}(0, \sigma_\varepsilon^2)$.

Then we have a correlation between observations at the same level equal to

$$\rho = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2}.$$

This ρ is known as the *intraclass correlation coefficient*.

Two-way ANOVA model

Two-way ANOVA model is appropriate for datasets that contain a continuous numerical response variable and two categorical predictors.

Y_{ijk} means the k^{th} measurement observed at the i^{th} $(k = 1, \cdots, n)$ level of the first factor $(i = 1, \cdots, I)$ and the j^{th} factor of the second factor $(j = 1, \cdots, J)$. With a balanced design, we have the additive model

$$Y_{ijk} = \mu_i + \nu_j + \epsilon_{ijk} = \mu + \tau_i + \alpha_j + \epsilon_{ijk},$$

where $\mu_i + \nu_j$ is the expected response within the $(i, j)^{th}$ level combination of the two factors, μ_i representing the effect on the expected response of the i^{th} level of the first factor and ν_j the effect of the j^{th} level of the second factor.

Two-way ANOVA model

The previous model assumes that the effects of the two factors are additive: the effect of the either factor is not changed depending on the level of the other factor at which the observations are being made.

No interaction between two factors!

Two sets of commonly used constraints:

- the "baseline" or "control group structure": $\tau_1 = \alpha_1 = 0$; or,
- the "grand mean" constraints: $\sum_{i=1}^{I} \tau_i = \sum_{j=1}^{J} \alpha_j = 0$.

Two-way ANOVA model

We still use the sequential F-statistic to do the hypothesis test. For example,

$$H_0 : \beta_{(2)} = 0,$$

$$F = \frac{SSR(\beta_{(2)} | \beta_{(1)}, \beta_0)/(J-1)}{MSE_{full}}$$

which has an F-distribution with $J-1$ numerator and $nIJ-(I+J-1)$ denominator degrees of freedom.

(in analogy to testing of a subset of βs in multiple linear regression)

ANCOVA models

A simple ANCOVA model with a continuous predictor x and a factor α:

$$Y_{ij} = \beta_0 + \alpha_i + \beta_1 x_{ij} + \varepsilon_{ij}.$$

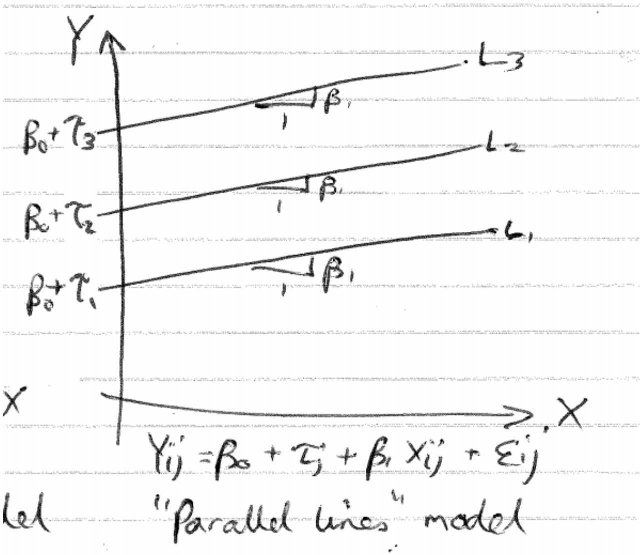
An ANCOVA model with a continuous predictor x, a factor α and also an interaction term:

$$Y_{ij} = \beta_0 + \alpha_i + \beta_1 x_{ij} + \gamma_i x_{ij} + \varepsilon_{ij}.$$

If we use indicator variables for the categorical predictor (e.g. "YES" and "NO"), we can have the following parameterisations:

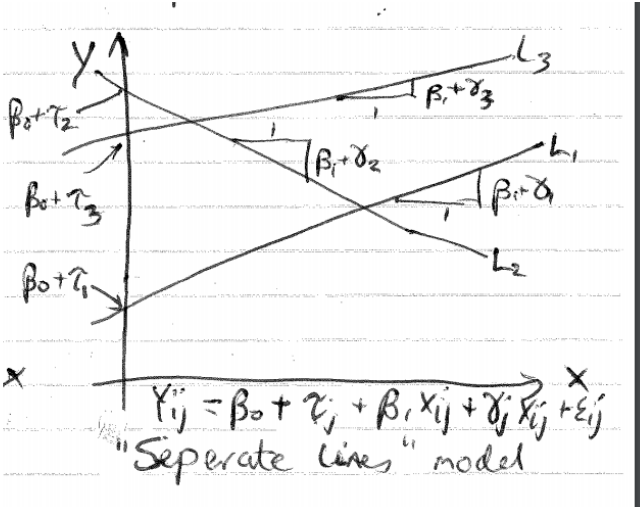
$$\hat{Y}_{ij} = \begin{cases} (\hat{\beta}_0 + \alpha_Y) + (\hat{\beta}_1 + \alpha_Y) X_{ij}, & \text{if YES} \\ \hat{\beta}_0 + \hat{\beta}_1 X_{ij}, & \text{if NO} \end{cases}$$

Diagram



Y
L_3
β_1
1
$\beta_0 + \tau_3$
L_2
β_1
1
$\beta_0 + \tau_2$
L_1
β_1
1
$\beta_0 + \tau_1$
X
$Y_{ij} = \beta_0 + \tau_j + \beta_1 X_{ij} + \varepsilon_{ij}$
"Parallel lines" model

Diagram



Y
X
L_3
L_1
L_2
$\beta_0+\tau_2$
$\beta_0+\tau_3$
$\beta_0+\tau_1$
$\beta_1+\gamma_3$
$\beta_1+\gamma_2$
$\beta_1+\gamma_1$
1
1
1
$Y_{ij} = \beta_0 + \tau_j + \beta_1 x_{ij} + \gamma_j x_{ij} + \varepsilon_{ij}$
"Seperate lines" model

Link functions

Apart from the canonical link functions, we have other commonly used link functions:

1. Logit: $g(p) = \log \frac{p}{1-p}$
2. Probit: $g(p) = \Phi^{-1}(p)$
3. Complementary log-log: $g(p) = \log(-\log(1-p))$

In GLM we will model $g(\boldsymbol{\mu}) = \hat{\boldsymbol{\eta}} = \mathbf{X}^T\boldsymbol{\beta}$, we need to do back transformation to get $\hat{\boldsymbol{\mu}}$.

Logistic regression model

The inverse of the logit function is called the logistic function (or inverse logit):

$$p = \frac{exp(\eta)}{1 + exp(\eta)}$$

Our logistic regression model for binary response is then:

$$g(p) = logit(p) = \log \frac{p}{1-p} = \beta_0 + \beta_1 X_1 + \cdots + \beta_q X_q$$

The response Y is assumed to have a Bernoulli distribution with probability p:

$$Y = \begin{cases} 1 & \text{with probability} \quad p \\ 0 & \text{with probability} \quad 1-p \end{cases}$$

Drop in deviance test

The Likelihood ratio test can be expressed as:

$$LRT = deviance_{reduced} - deviance_{full}$$

Deviance values can be found in summary outputs. We still compare the drop-in-deviance result to a χ^2_d distribution, with d denoting the difference in the number of parameters.

Delta Method

The delta method is a statistical approach to derive an approximate probability distribution for a function of an asymptotically normal estimator using the Taylor series approximation.

If a sequence of random variables $Y_1, \cdots, Y_n$ satisfying

$$\sqrt{n}(Y_i - \theta) \xrightarrow{D} \mathcal{N}(0, \sigma^2),$$

where θ and σ^2 are finite valued constants, then

$$\sqrt{n}\,(g(Y_i) - g(\theta)) \xrightarrow{D} \mathcal{N}(0, [g'(\theta)]^2\sigma^2).$$

Confidence interval for $g^{-1}(X^T\boldsymbol{\beta})$

When we want to calculate a 95% confidence interval for a function of the parameters $\boldsymbol{\beta}$, say $\mu = g^{-1}(X^T\boldsymbol{\beta})$, we can firstly compute a confidence interval for $X^T\boldsymbol{\beta}$ as $\{L, U\}$, and then apply the function $g^{-1}()$ to both bounds L and U.

The desired confidence interval is given by $\{g^{-1}(L), g^{-1}(U)\}$.

(Proof is on Page 41 of the lecture brick on Wattle.)

Deviance

The *deviance* or *residual deviance*, $D(\hat{Y}, Y)$ is defined as

$$D(\hat{Y}, Y) = 2\phi\{\ell(Y, \phi) - \ell(\hat{Y}, \phi)\},$$

which measures the (scaled) difference between the log-likelihood for the **observed data** and the log-likelihood of the the **fitted values**, and thus small values of the deviance indicate that a model fits the observed data well.

Scaled deviance

For independent observations Y_i and exponential family errors, we have

$$D(\hat{Y}, Y) = 2\sum_{i=1}^{n}\{Y_i(\hat{\theta}_{saturated} - \hat{\theta}) - b(\hat{\theta}_{saturated}) + b(\hat{\theta})\}.$$

(exponential family and $b(\cdot)$ functions on Page 33 of the brick)

Then we have can write likelihood ratio statistics for comparison between a saturated model and the model of interest as

$$\text{Likelihood ratio} = D^* = \frac{D(\hat{Y}, Y)}{\phi}$$

Dispersion

The dispersion parameter ϕ indicates if we have more or less than the expected variance. We have already seen that $\phi = 1$ for Binomial and Poisson distributions. In the **summary** output we have **dispersion** parameter defined as

$$\phi_{assumed} = \begin{cases} MSE = \frac{1}{n-p}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2, & \text{Normal} \\ 1, & \text{Binomial and Poisson} \\ CV = \frac{1}{n-p}\sum_{i=1}^{n}(\frac{Y_i - \hat{Y}_i}{\hat{Y}_i})^2, & \text{Gamma} \end{cases}$$

where CV is the estimated coefficient of variation (relative standard deviation) for the gamma distribution.

Alternative estimates of dispersion

An alternative estimate of ϕ for all GLMs is

$$\phi_{alt} = \frac{D(\hat{Y}, Y)}{n - p}.$$

If $\phi_{alt} = \phi_{assumed} \longrightarrow$ model is “good”.
If $\phi_{alt} < \phi_{assumed} \longrightarrow$ model is **under-dispersed**.
If $\phi_{alt} > \phi_{assumed} \longrightarrow$ model is **over-dispersed**.

Goodness of fit test

We can also use deviance to assess model fit.

$$\frac{D(\hat{Y}, Y)}{\phi} \sim \chi^2_{n-p} \quad \text{under } H_0$$

Pearson residual

If we define $e_i = Y_i - \hat{Y}_i$ as residual, e_i for a GLM does not behave quite nicely as in SLR models. In particular, we know that the variance of the Y_i's is not constant, but is instead proportional to the variance function $V(\mu_i)$. Thus, even if the chosen model is correct, the residuals will not display a homoscedastic spread.

Pearson residual as

$$r_i = \frac{e_i}{\sqrt{V(\hat{Y}_i)}}$$

```
residuals(model, "pearson")
```

(see "R Example: Residual Plots" on Wattle)

Deviance residual

Deviance residual d_i is defined based on $D(\hat{Y}, Y)$, so that

$$\sum_{i=1}^{n} d_i^2 = D(\hat{Y}, Y).$$

```
residuals(model, "deviance")
```

(see "R Example: Residual Plots" on Wattle)

Studentised residuals

- $Var(r_i) \approx Var(d_i) \approx \hat{\phi}(1 - h_{ii})$
- Since the variance are approximate values, these Studentised residuals are rarely used in this course.
- Example in “R Example: Residual Plots” and Page 64 of the brick.

```
residuals(model)/sqrt(summary(model)$dispersion*
(1-influence(model)$hat))
```

Deletion residuals

- $r_i^* = \frac{Y_i - \hat{Y}_{i,-i}}{\sqrt{V(\hat{Y}_{i,-i})}}$
- Similar idea to the so-called PRESS residuals of multiple linear regression
- Relevant code in "R Example: Outliers"
- Used for assessing outliers

Two tests

There are two "classical" tests of independence:

To test H_0: no association between Factor 1 and Factor 2

- Likelihood ratio

$$2\sum_{i=1}^{R}\sum_{j=1}^{C} O_{ij} \log\left(\frac{O_{ij}}{E_{ij}}\right)$$

- Pearson Chi-squared

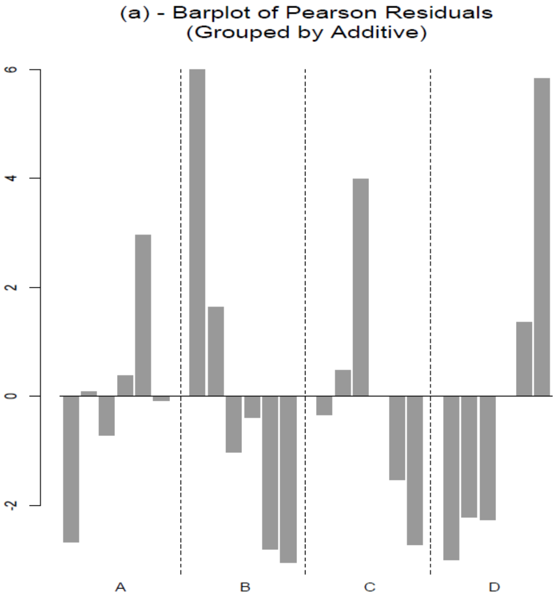
$$\sum_{i=1}^{R}\sum_{j=1}^{C} \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

Both have an asymptotic χ^2 distribution with $(r-1)(c-1)$ degrees of freedom

Interpretation of Pearson residuals

Interpretation of Pearson residuals: (Page 78 of brick)

- $(i, j)^{\text{th}}$ cell with a large positive residual $\longrightarrow O_{ij} \gg E_{ij}$
- E_{ij} is the expected value under the independence assumption
- It indicates that individuals in the j^{th} column are more likely to be in the i^{th} row than individuals in the other columns
- vice versa


(a) - Barplot of Pearson Residuals
(Grouped by Additive)
6
4
2
0
-2
A
B
C
D