STATS 205: Hierarchical Linear Models **Fall 2017**

Lecture 4: Linear Models with Categorical Data

Lecturer: Prof. Jingyi Jessica Li *Subscribers: Zheqi Wu and Ziyi Jiang*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

Previously scribers: Ruochen Jiang and Zhanhao Peng

4.1 Recap

4.1.1 Fisher's z-transformation

population (Pearson) correlation

$$\rho = \frac{Cov(\mathbf{X}, \mathbf{Y})}{\sqrt{Var(\mathbf{X})Var(\mathbf{Y})}}$$

sample (Pearson) correlation

$$r_{xy} = \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2 \sum\limits_{i=1}^{n}(y_i - \bar{y})^2}}$$

Hypothesis about the value of the population correlation coefficient ρ between variables $\mathbf{X}$ and $\mathbf{Y}$ can be tested using the Fisher transformation applied to the sample correlation coefficient. We know that

$$\sqrt{n-3}\left(z - \frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right)\right) \xrightarrow{\text{d}} N(0,1),$$

where

$$z = \frac{1}{2}\log\left(\frac{1+r}{1-r}\right)$$

To test the null hypothesis $H_0 : \rho = 0$, $H_1 : \rho \neq 0$

$$z \overset{approx}{\sim} N\left(0, \frac{1}{n-3}\right)$$

4.2 One-way ANOVA for Categorical Predictors

- One categorical predictor (i.e., factor) with I level
- n_i: number of observations in the i^{th} level
- $\sum_{i=1}^{I} n_i = n$

- Y_{ij}: the j^{th} response in the i^{th} level, j = 1, ..., n_i
- random structure: $Y_{ij} \sim N(\mu_i, \sigma^2)$, $i = 1, \ldots, I$; $j = 1, \ldots, n_i$
- systematic structure: $\mu_i = \mu + \alpha_i$, $i = 1, \ldots, I$
- In order to guarantee identifiability: α_1=0.
- We have I parameters: μ, $\alpha_2, \ldots, \alpha_k$. Where μ is the intercept and α_i describes the expected difference between level i and level 1.

4.2.1 estimators

$$\begin{aligned}
\hat{\mu} &= \bar{Y}_{1\cdot} \\
\hat{\alpha}_2 &= \hat{\mu}_2 - \hat{\mu} = \bar{Y}_{2\cdot} - \bar{Y}_{1\cdot} \\
&\vdots \\
\hat{\alpha}_I &= \hat{\mu}_I - \hat{\mu} = \bar{Y}_{I\cdot} - \bar{Y}_{1\cdot}
\end{aligned}$$

Notice here $\bar{Y}_{i\cdot}$ denotes the average response in level i.

We write this in terms of Y = $X\beta + \epsilon$ with dimensions: Y: $n \times 1$, X: $n \times I$, β: $I \times 1$, ϵ: $n \times 1$, and $\epsilon \sim N(0, \sigma^2 I_n)$

$$E(Y) = \underset{n\times 1}{\begin{bmatrix} E(Y_{11}) \\ \vdots \\ E(Y_{1n_1}) \\ \vdots \\ E(Y_{I1}) \\ \vdots \\ E(Y_{In_I}) \end{bmatrix}} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_1 \\ \vdots \\ \mu_I \\ \vdots \\ \mu_I \end{bmatrix} = \begin{bmatrix} \mu \\ \vdots \\ \mu \\ \vdots \\ \mu + \alpha_I \\ \vdots \\ \mu + \alpha_I \end{bmatrix} = \overset{X}{\underset{n\times I}{\begin{pmatrix} 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \ldots & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \ldots & 0 & 1 \end{pmatrix}}} \underset{I\times 1}{\begin{bmatrix} \mu \\ \alpha_2 \\ \vdots \\ \alpha_I \end{bmatrix}}$$

$$\underset{I\times 1}{\beta} = \begin{bmatrix} \mu \\ \alpha_2 \\ \vdots \\ \alpha_I \end{bmatrix} \text{ and } \underset{I\times 1}{\hat{\beta}} = (X^T X)^{-1} X^T Y,\ Var(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$$

4.2.2 Hypothesis Test

(1) H_0: $\alpha_i = 0$, $i = 2, \ldots, I$. We will use t-test.

(2) Let $\alpha = (\alpha_2, \ldots, \alpha_I)^T = (\alpha^{(1)}, \alpha^{(2)})^T$
H_0: $\alpha^{(2)} = 0$. We will use Wald test or likelihood ratio test.

(3) H_0: $\alpha_2 = \alpha_3 = \ldots = \alpha_k = 0$ (It's a special case of (2))
We will use the one-way ANOVA.

SV	SS	DF	MS	F
X	SSR	$I-1$	$MSR = \frac{SSR}{I-1}$	$\frac{MSR}{MSE}$
Residual	SSE	$n-I$	$MSE = \frac{SSE}{n-I}$	
total	SST	$n-1$		

If we let n $\to \infty$, $(I-1) \cdot F \sim \chi^2_{(n-I)}$

4.2.3 Point biserial correlation

Point biserial correlation is the correlation between categorical variable X and continuous variable Y.

$$r^2_{pb} = \frac{SSR}{SST} \in [0,1]$$

4.2.4 Discretized continuous model

If you have a continuous predictor, you may consider discretizing it and use one-way ANOVA.

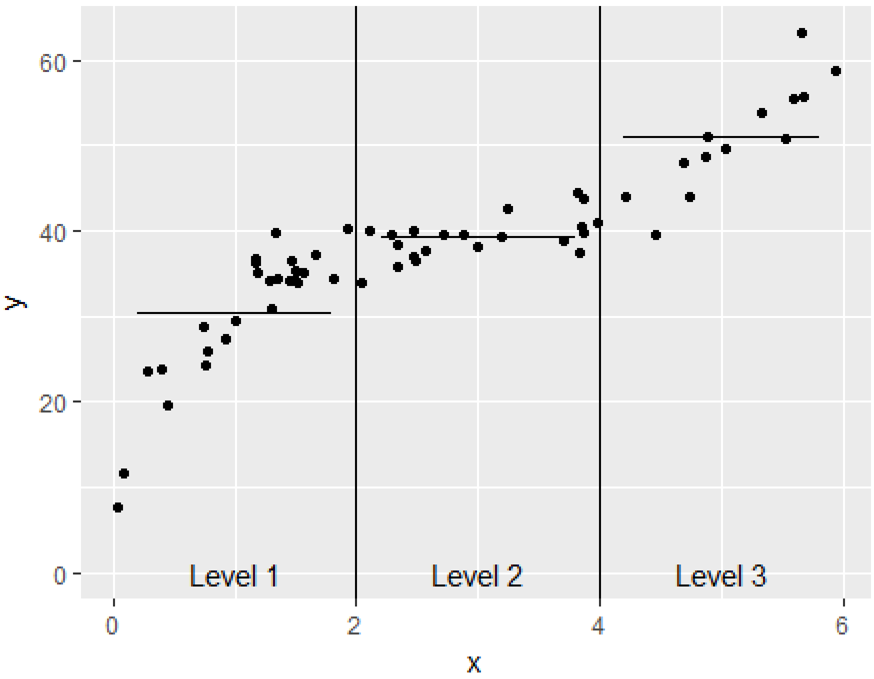

Figure 4.1: *Regression when splitting up x into three categories. The fitted value for $\hat{y}$ is the mean of y at each category. This is an example of a nonparametric regression.*

The simple linear regression model has 2 parameters, β_0 and β_1, corresponding to the intercept and the slope.

The one-way ANOVA model has 3 parameters μ, α_2, and α_3, each representing the mean of y in levels 1, 2 and

3 of x. The vertical bars represent the different levels of x on the nominal scale. The horizontal bars represent the mean of y within a level. In this aspect, one-way ANOVA is a special type of nonparametric/nonlinear regression. Thus this model is more complex comparing to the simple linear model.

4.3 Two-way ANOVA (Without Interaction Effect)

- One-way ANOVA: one categorical (factor) predictor
- Two-way ANOVA: two categorical (factor) predictors

We are going to talk about Two-way ANOVA that has:

- I levels of factor 1
- J levels of factor 2
- n_{ij} observations in level i of factor 1 and level j of factor 2
- $n = \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij}$ is the total number of observations

Random structure

$$Y_{ijk} \sim N(\mu_{ij}, \sigma^2),$$

the distribution of the k^{th} observation in level (i, j) for $k \in \{1, 2, \ldots, n_{ij}\}$

Systematic structure (additive model — in general, we may apply some non-linear transformation to the predictor, but add them all. And the effect of α_i has no influence on effect of γ_j)

$$\mu_{ij} = \mu + \alpha_i + \gamma_j,$$

where

- μ: constant
- α_i: effect of level i of factor 1
- γ_j: effect of level j of factor 2

Table 4.1: Mean in each case

	$F1_1$	$F1_2$	$\ldots$	$F1_I$
$F2_1$	$\mu + \alpha_1 + \gamma_1$	$\mu + \alpha_2 + \gamma_1$	$\ldots$	$\mu + \alpha_I + \gamma_1$
$F2_2$	$\mu + \alpha_1 + \gamma_2$	$\mu + \alpha_2 + \gamma_2$	$\ddots$	$\mu + \alpha_I + \gamma_2$
$\vdots$	$\vdots$	$\ddots$	$\ddots$	$\vdots$
$F2_J$	$\mu + \alpha_1 + \gamma_J$	$\mu + \alpha_2 + \gamma_J$	$\ldots$	$\mu + \alpha_I + \gamma_J$

In order to make the model identifiable, we assume that:

- $\alpha_1 = \gamma_1 = 0$
- The effect of level i of factor 1 does not depend on the level j of factor 2 for all i and j.

Then, the design matrix X will be

$$X = \begin{array}{c} \begin{matrix} 1 & \alpha_2 & \alpha_3 & \dots & \alpha_I & \gamma_2 & \gamma_3 & \dots & \gamma_J \end{matrix} \\ \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 & 1 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 1 & \dots & 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 1 \end{pmatrix} \end{array}$$

Then,

$$\underset{(I+J-1)\times 1}{\hat{\beta}} = (X^T X)^{-1} X^T Y = \begin{bmatrix} \hat{\mu} \\ \hat{\alpha}_2 \\ \hat{\alpha}_3 \\ \vdots \\ \hat{\alpha}_I \\ \hat{\gamma}_2 \\ \hat{\gamma}_3 \\ \vdots \\ \hat{\gamma}_J \end{bmatrix}$$

After that, we can do t test, Wald test and ANOVA test as needed.

Table 4.2: ANOVA

SV	SS	DF	MS	F
Factor1	SSR_1	$I-1$	MSR_1	
Factor2 | Factor1	$SSR_{2\|1}$	$J-1$	$MSR_{2\|1}$	$F = \frac{MSR_{2\|1}}{MSE}$
Residual	SSE	$N-(I+J-1)$	MSE	

$$SSR_1 + SSR_{2|1} = SSR = SSR_2 + SSR_{1|2}$$

The F statistic above is for testing whether the net effect of factor 2 is zero in the model with both factors. If we want to test whether the gross effect of factor 2 is zero, we should use one-way ANOVA by including factor 2 only.

- Gross Effect: *unadjusted effect of factor 2 (Also called marginal effect which can be get by using factor 2 as the only predictor)*
- Net effect: *adjusted effect (factor 2 | other factors) or conditional effect/additional effect*
- To detect Net effect using *ANOVA* function in R, we always put the predictor we are interested as the last predictor, eg. if we want to study the net effect of F_2, we use anova(lm($Y \sim F_1 + F_2$))

Estimate of the unadjusted effect of level j of factor 2 on observations: $\hat{\mu}_j = \bar{Y}_{.j.}$ (One-Way ANOVA)

Estimate of the adjusted effect of level j of factor 2 on observations: $\hat{\mu}_{.j} = \frac{1}{n}\sum_{i=1}^{I}(\sum_{j=1}^{J} n_{ij})\hat{\mu}_{ij} = \frac{1}{n}\sum_{i=1}^{I}(\sum_{j=1}^{J} n_{ij})\bar{Y}_{ij.}$ (Two-way ANOVA)

4.4 Two-way ANOVA with Interaction Effects

- Systematic structure: $\mu_{ij} = \mu + \alpha_i + \gamma_j + \eta_{ij}$
  where η_{ij} is the notation for interaction where $i = 1, \ldots, I$ and $j = 1, \ldots, J$
- Identifiability: $\alpha_1 = \gamma_1 = \eta_{1j} = \eta_{i1} = 0$

$$\beta = \begin{bmatrix} \mu \\ \alpha_2 \\ \vdots \\ \alpha_I \\ \gamma_2 \\ \vdots \\ \gamma_J \\ \eta_{22} \\ \vdots \\ \eta_{IJ} \end{bmatrix}_{(m \times n) \times 1}$$

The X matrix will be

$$\begin{array}{c} \begin{array}{ccccccccccccc} 1 & \alpha_2 & \alpha_3 & \ldots & \alpha_I & \gamma_2 & \gamma_3 & \ldots & \gamma_J & \eta_{22} & \eta_{23} & \ldots & \eta_{IJ} \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 & 0 & 1 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ 1 & 1 & 0 & \ldots & 0 & 1 & 0 & \ldots & 0 & 1 & 0 & \ldots & 0 \\ 1 & 1 & 0 & \ldots & 0 & 0 & 1 & \ldots & 1 & 0 & 1 & \ldots & 0 \\ 1 & 0 & 1 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ 1 & 0 & 1 & \ldots & 0 & 1 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \ldots & 1 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ 1 & 0 & 0 & \ldots & 1 & 0 & 0 & \ldots & 1 & 0 & 0 & \ldots & 1 \end{pmatrix} \end{array}$$

The total sum of squares can now be partitioned into four sources:

- Factor 1
- Factor 2 $\|$ Factor 1
- Interaction
- Error

4.5 Analysis of Covariance Models

- Combination of categorical factors and continuous variables.
- x continuous with 1 degree of freedom, z categorical with I levels and $I - 1$ degrees of freedom.
- n_i observations in level i of z.
- $n = \sum_{i=1}^{I} n_i$

- Random structure: $Y_{ij} \sim N(\mu_{ij}, \sigma^2), j = 1, ..., n_i$
- Systematic structure: $\mu_{ij} = \mu + \alpha_i + \gamma x_{ij}$. Impose $\alpha_1 = 0$ for identifiability.

Then this model represents I parallel lines, one for each group. The X matrix will look like

$$\begin{array}{c} \begin{array}{cccccc} 1 & \alpha_2 & \alpha_3 & \dots & \alpha_I & \gamma \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & x_{11} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \dots & 0 & x_{1n_1} \\ 1 & 1 & 0 & \dots & 0 & x_{21} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & \dots & 0 & x_{2n_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \dots & 1 & x_{I1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \dots & 1 & x_{In_I} \end{pmatrix} \end{array}$$

With total degrees of freedom $n - 1$, x continuous with 1 degrees of freedom, z categorical with $I - 1$ degrees of freedom, and residual degrees of freedom $n - (I + 1)$.

We can drop the parallel lines assumption. Then $\mu_{ij} = \mu + \alpha_i + (\gamma + \eta_i)x_{ij}$.

- Idenfiability conditions: $\alpha_1 = \eta_1 = 0$
- Design matrix X? Homework question.
- Can test $H_0 : \eta_2 = \cdots \eta_k = 0$ by Wald or LRT.