χ^2 test:

H_0: Observed distribution is compatible with a given distribution. [Multinomial]

O: Observed counts.

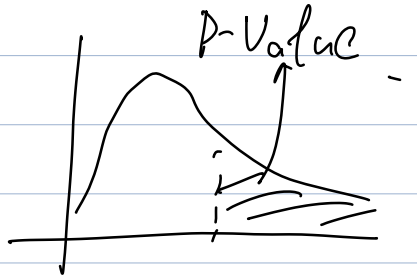
E: Expected counts.

K: Number of categories

n: Total number of outcomes.

$\chi^2 = \sum_{i=1}^{K} (O_i - E_i)^2 / E_i$

Under H_0: $\chi^2 \sim \chi^2_{k-1}$



One Sample t-test

$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$

$(\mu > \mu_0)$

$(\mu < \mu_0)$

$T = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$ with $s = \sqrt{s^2}$ ← Sample Variance.

$\bar{x}$: Sample mean $T \sim$ student t with $(n-1)$ df.

Test for 2 means:

1) Independent samples:

a) Unequal variances

Assumptions: — Both are simple random samples

— $n_1 > 30$, $n_2 > 30$ or both samples are from Normal distribution.

$H_0: \mu_1 - \mu_2 = \mu_0 \qquad H_1: \mu_1 - \mu_2 \neq \mu_0$

Generally, $\mu_0 = 0$

$$T = \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

$$df = \frac{(A+B)^2}{\frac{A^2}{n_1-1} + \frac{B^2}{n_2-1}} \qquad A = s_1^2/n_1 \qquad B = s_2^2/n_2$$

b) $\sigma_1^2 = \sigma_2^2$

$$T = \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{\sqrt{s_p^2/n_1 + s_p^2/n_2}}$$

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1-1)+(n_2-1)} \qquad T \sim t(n_1+n_2-2)$$

2) Matched Pairs

Assumptions: - Sample of match pairs
- n: number of pairs. $n \geq 30$ or differences come from a Normal distribution.

$H_0: \mu_d = 0 \qquad H_1: \mu_d \neq 0$

$$T = \frac{\bar{d}}{s_d/\sqrt{n}} \qquad t(n-1)\ df$$

Density Estimation:

$$\hat{f}(x) = \frac{1}{nb} \sum_{j=1}^{n} K\left(\frac{x - x_j}{b}\right) \quad ?$$

Defaut: $\hat{b} = 0.9 \min(\hat{\sigma}, R/1.34)\, n^{-1/5}$

Test for proportion:

$H_0: \hat{p} = p$

Page: Data Explanation. challenges. references.

Oct. 14.

Simple linear Regression:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i. \quad i = 1, \cdots, n.$$

$\varepsilon_i \overset{i.i.d.}{\sim} N(0, \sigma^2)$

Least Square Estimation: → Don't need any assumption

Finding $\hat{\beta}_0$ and $\hat{\beta}_1$ such that the minimize:

$$\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 = f(\beta_0, \beta_1)$$

$$\frac{\partial f(\beta_0, \beta_1)}{\partial \beta_0} = \sum_{i=1}^{n} -2(y_i - \beta_0 - \beta_1 x_i) = 0 \qquad ①$$

$$\frac{\partial f(\beta_0, \beta_1)}{\partial \beta_1} = \sum_{i=1}^{n} -2x_i(y_i - x_i\beta_1 - \beta_0) = 0 \quad ②$$

Using ①: $\sum_{i=1}^{n} y_i - n\beta_0 - \beta_1 \sum_{i=1}^{n} x_i = 0$

$$\Rightarrow \bar{y} - \beta_0 - \beta_1 \bar{x} = 0$$

$$\Rightarrow \bar{y} = \beta_0 + \beta_1 \bar{x} \quad ★$$

Using ②:

$$\sum_{i=1}^{n} x_i y_i - x_i^2 \beta_1 - x_i \beta_0 = 0 \quad ★★$$

$$\Rightarrow \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i^2 \cdot \beta_1 - \sum_{i=1}^{n} x_i \beta_0 = 0$$

By solving ★ and ★★:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} \qquad \beta_0 = \bar{y} - \bar{x}\beta_1$$

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} \Rightarrow \begin{cases} SS_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 \\ SS_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \\ SS_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2 \end{cases}$$

$\hat{y}_i = \hat{\beta}_0 + x_i \hat{\beta}_1 \qquad \varepsilon_i = y_i - \hat{\beta}_0 - x_i \hat{\beta}_1$

$SSE = \sum_{i=1}^{n} (y - \hat{y}_i)^2$

$MSE = SSE / n-2 = \hat{\sigma}^2$

$E(\hat{\sigma}^2) = \sigma^2$

Hypothesis Testing.

$H_0: \beta_1 = 0 \qquad H_1: \beta_1 \neq 0$

$$T = \frac{\hat{\beta}_1 - 0}{\frac{\hat{\sigma}^2}{\sqrt{SS_{xx}}}}$$

$T \sim$ student-t with $n-2$ df.

* Confidence interval for β_1: at α level of significance:

The $(1-\alpha)\%$ CI (L, U) can be obtained with:

$$L = \hat{\beta}_1 - t_{\alpha/2, n-2} \frac{\hat{\sigma}^2}{\sqrt{SS_{xx}}}$$

$$U = \hat{\beta}_1 + t_{\alpha/2, n-2} \frac{\hat{\sigma}^2}{\sqrt{SS_{xx}}}$$

$$\hat{\sigma}^2 = \frac{SS_E}{n-2}.$$

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \rightarrow SSR}{\sum_{i=1}^{n} (y_i - \bar{y})^2 \rightarrow SST}$$

Confidence Interval for the mean response:

Interval for x^*:

$$\mu_{y|x^*} = \beta_0 + \beta_1 x^*$$

$$\hat{\mu}_{y|x^*} = \hat{\beta}_0 + \hat{\beta}_1 x^*$$

Confidence interval: $(1-\alpha)100\%$

$$\hat{\mu}_{y|x^*} \pm t_{\alpha/2, n-2} \cdot se\,\hat{\mu}_{y|x^*}$$

$$se\,\hat{\mu}_{y|x^*} = \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

Prediction interval: prediction for a new observation at x^*

$$y^{new} = \beta_0 + \beta_1 x^* + \varepsilon^{new}$$

Point estimation is the same.

$$\hat{y}_{new} = \hat{\beta}_0 + \hat{\beta}_1 x^*$$

But the interval is different:

$$\hat{y}_{new} \pm t_{\alpha/2,\, n-2} \cdot Se\,\hat{y}_{new}$$

$$Se\,\hat{y}_{new} = \hat{\sigma}\sqrt{1 + 1/n + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

H_0 for F test:

$$\beta_1 = \beta_2 = \cdots = 0$$

H_1 for F test:

At least 1 of β_i is not 0.

Multiple Regression:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} + \varepsilon_i$$

$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ $\quad i = 1 : n$

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix} \quad X = \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{p1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{pn} \end{pmatrix}$$

$\dim(X) = n \times (p+1)$ $\quad \dim(\beta) = (p+1) \times 1$

$\dim(y) = n \times 1$ $\quad \dim(\varepsilon) = n \times 1$ $\quad \varepsilon \sim N(0, \sigma^2 \cdot I_n)$

$$y = X\beta + \varepsilon \quad , \quad \varepsilon \sim N(0, \sigma^2 I_n)$$

1) LSE : Find $\hat{\beta}$ that minimizes:

$$S(\beta) = \| y - X\beta \| = (y - X\beta)^T (y - X\beta)$$

2) MLE:

$$y \sim N(X\beta, \sigma^2 \cdot I_n)$$

$$L(\beta, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(y_i - \beta_0 - \beta_1 x_{1i} - \cdots - \beta_p x_{pi})^2}{2\sigma^2}\right)$$

or, equivalently:

$$L(\beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{(y - X\beta)^T (y - X\beta)}{2\sigma^2}\right\}$$

Maximizing $L(\beta, \sigma^2)$ is the same as maximizing:

$$\log L(\beta, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{(y - X\beta)^T (y - X\beta)}{2\sigma^2}$$

For method 1:

LSE: we find $\frac{\partial S(\beta)}{\partial \beta_0} = 0, \ldots, \frac{\partial S(\beta)}{\partial \beta_p} = 0$

If X is a full rank matrix with $n \geq p+1$, and rank $= p+1$. Then the solution is unique and given by:

$$(X^T X)^{-1} X^T Y.$$

For method 2:

MLE: we find: $\frac{\partial l(\beta, \sigma^2)}{\partial \beta} = 0$ | $\frac{\partial l(\beta, \sigma^2)}{\partial \sigma^2} = 0$ (if σ^2 not known)

If σ^2 known: The MLE with X a full rank matrix and $n \geq p+1$ is also $(X^T X)^{-1} X^T Y$.

But we also have σ^2:

so, we also find $\frac{\partial l(\beta, \sigma^2)}{\partial \sigma^2} = 0$. not unbiased

This gives us:

$$\hat{\sigma}^2_{MLE} = \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{n} = \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n}$$

So, we use: $\hat{\sigma}^2 = \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n-(p+1)}$

Properties of $\hat{\beta}$:

$$E(\hat{\beta}) = E((X^TX)^{-1}X^TY)$$
$$= (X^TX)^{-1}X^TX\beta$$
$$= \beta$$

$$Var(\hat{\beta}) = (X^TX)^{-1}X^T Var(y) X (X^TX)^{-1}$$
$$= (X^TX)^{-1}X^T \sigma^2 \cdot I_n X (X^TX)^{-1}$$
$$= \sigma^2 (X^TX)^{-1}$$

$\hat{\beta} \sim N(\beta, \sigma^2(X^TX)^{-1})$ $\quad \hat{\beta}_i \sim N(\beta_i, \sigma^2 (X^TX^{-1})_{ii})$

F-test for Nested Model:

M_1: $y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} + \varepsilon_i$

M_2: $y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi} + \beta_{p+1} x_{p+1,i} + \cdots + \beta_{p+q} x_{p+q,i}$

H_0: $\beta_{p+1} = \beta_{p+2} = \cdots = \beta_{p+q} = 0$ vs: $+\varepsilon_i$

H_a: At least 1 of $\beta_{p+1} \cdots \beta_{p+q}$ is not 0.

Fit M_1 and M_2 and get:

$\frac{SSE_1}{X_1}$ and $\frac{SSE_2}{X_2}$.

$\hat{\beta}^{(1)}$ $\quad \hat{\beta}^{(2)}$

$SSE_1 = (y - X_1\hat{\beta}^{(1)})^T (y - X_1^T \hat{\beta}^{(1)})$

$SSE_2 = (y - X_2\hat{\beta}^{(2)})^T (y - X_2^T \hat{\beta}^{(2)})$

Note that: $SSE_1 > SSE_2$.

$$F = \frac{(SSE_1 - SSE_2)/q_0}{SSE_2/(n-(p+q_0+1))}$$

Under H_0: $F \sim$ F-distribution with d.f. of q_0 in the numerator, $n-(p+q_0+1)$ in the denominator.

What about comparing models that are not nested:

- AIC: Akaike's information

If you have a model M: ↗ # of parameter in the model

$$AIC(M) = -2\log \mathcal{L}(M) + 2p(M)$$

↳ likelihood of model M evaluated at the MLE.

- BIC: Bayesian Information Criterion.

$$BIC(M) = -2\log(\mathcal{L}(M)) + \log(n)\cdot p(M)$$

↳ # of observations.

If you have models $M_1, M_2 \cdots M_k$, chose the one that minimizes AIC or BIC.

Adjusted R^2:

$$R^2_{adj} = 1 - \left[\frac{SSE}{n-(p+1)}\right] / \left[\frac{\Sigma(y_i-\bar{y})^2}{n-1}\right]$$

$$SST = \Sigma(y_i-\bar{y})^2$$

Oct. 21: One-Way ANOVA.

$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ $\quad \varepsilon_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2)$

$i = 1 : a$. $\quad j = 1 : n_i$

↳ If n_j are same, the model is called balanced model.

$n_i = n$ for all i

1 factor with a levels. can we write this model as

$y = X\beta + \varepsilon \quad \varepsilon \sim N(0, \sigma^2 I)$.

$$y = \begin{pmatrix} y_{11} \\ \vdots \\ y_{1,n_1} \\ y_{21} \\ \vdots \\ y_{2,n_2} \\ \vdots \\ y_{a1} \\ \vdots \\ y_{a,na} \end{pmatrix} \quad \beta = \begin{pmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_a \end{pmatrix} \quad \varepsilon = \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1,n_1} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{2,n_2} \\ \vdots \\ \varepsilon_{a1} \\ \vdots \\ \varepsilon_{a,na} \end{pmatrix}$$

$$X = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & 0 \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 1 & 0 \cdots & 0 \\ \vdots & \vdots & \vdots & & \ddots \\ 1 & 0 & \cdots & \ddots & 0 & 1 \\ \vdots & \vdots & & & \vdots \\ 1 & 0 & \cdots & \ddots & 0 & 1 \end{pmatrix}$$

→ Group 1

→ Group 2

→ Group

X is not full rank.

↳ If X is not full rank, MLE is not unique

Add restrictions can solve this problem, or look at different parameterizations of the model.

(a) Consider the paramet:

$$y_{ij} = \mu_i + \varepsilon_{ij} \quad \begin{cases} i = 1:a \quad j = 1:n_i \\ \varepsilon_{ij} \overset{i.i.d}{\sim} N(0, \sigma^2) \end{cases}$$

$$\beta = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_a \end{pmatrix}$$

The resulting X is a full rank.

$$X = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ 1 & \cdots & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \\ \vdots & & & \vdots \\ 0 & \cdots & 0 & 1 \end{pmatrix} \begin{matrix} \\ \\ \rightarrow \text{Group } 1 \\ \\ \\ \\ \rightarrow \text{Group } a. \end{matrix}$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

$$X^T X = \begin{pmatrix} n_1 & & & 0 \\ & n_2 & & \\ & & \ddots & \\ 0 & & & n_a \end{pmatrix} \quad (a \times a)$$

$$(X^T X)^{-1} = \begin{pmatrix} 1/n_1 & & & 0 \\ & 1/n_2 & & \\ & & \ddots & \\ 0 & & & 1/n_a \end{pmatrix}$$

$$X^T Y = \begin{pmatrix} \sum_{j=1}^{n_1} y_{1,j} \\ \vdots \\ \sum_{j=1}^{n_a} y_{a,j} \end{pmatrix}$$

$$\hat{\beta} = \begin{pmatrix} \bar{y}_{1\cdot} \\ \vdots \\ \bar{y}_{a\cdot} \end{pmatrix} \qquad \hat{\mu}_i = \bar{y}_{i\cdot} = \sum_{j=1}^{n_i} y_{i,j}$$

Hypothesis Testing:

$H_0: \mu_1 = \mu_2 = \cdots = \mu_a$ vs H_1: at least one μ_i is different from the rest

7-test If we fail to reject, then we are

$y_{i,j} = \mu + \varepsilon_{i,j}$ $\varepsilon_{i,j}$ i.i.d. $N(0, \sigma^2)$

$\hat{\mu} = \bar{y}_{..}$ 7-test between n μ an 1 μ.

b) Add one restriction to this model.

$y_{i,j} = \mu + \alpha_i + \varepsilon_i$ $\varepsilon_{i,j}$ i.i.d. $N(0, \sigma^2)$

possible restrictions:

i) $\alpha_{i_0} = 0$ for one i_0.

ii) $\sum_{i=1}^{a} \alpha_i = 0$

$n-1$ All Vars

$k-1$ All

↓

$k \cdot r$

Test for 2 groups:

$H_0: \mu_i = \mu_j$ vs $H_1: \mu_i \neq \mu_j$

$$T = \frac{\bar{y}_i - \bar{y}_j}{\sqrt{MSZ(1/n_i + 1/n_j)}} \qquad df = N - a.$$

↳ MSZ for linear regression.

Threshold is:

$$T_{\alpha/2,\, N-a} \cdot \sqrt{MSZ(1/n_i + 1/n_j)}.$$

α = P(Type 1 error)

= P(Reject | H_0 = T)

= P(Incorrect significant result)

Pr(At least 1 significant is incorrect)

= 1 - P(No significant result) = $1 - (1-\alpha)^k$

If $k = 20$, $\alpha = 0.05$

$1-(1-\alpha)^k \approx 0.64$

Oct. 26.

Randomised Block Models:

more than 2-dimensional paired t-test.

Multiple Block Factors:

Example: Latin square design.

	Z_1	Z_2	Z_3	Z_4
R_1	A	B	C	D
R_2	B	C	D	A
R_3	C	D	A	B
R_4	D	A	B	C

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk}$$

α_i: Block 1, β_j: Block 2, γ_k: Factor.

$i = 1:g$ $\quad j = 1:g$ $\quad k = 1:g$.

# of parameters: $1 + (g-1) + (g-1) + (g-1) = 3g - 2$

# of Observations: g^2

degree of freedom: $g^2 - 3g - 2$

Two-Way ANOVA Model:

Full model with interactions:

$$y_{i,j,k} = \mu + \alpha_i + \beta_j + \gamma_{i,j} + \varepsilon_{i,j,k} \qquad \varepsilon_{i,j,k} \overset{i.i.d.}{\sim} N(0, \sigma^2)$$

$i = 1:a \qquad j = 1:b \qquad k = 1:n_{i,j}$

Model with Interactions

$$y_{i,j,k} = \mu + \alpha_i + \beta_j + \varepsilon_{i,j,k} \quad (\text{no interactions})$$

$$y = X\beta + \varepsilon \qquad \hat{\beta} = (X^T X)^{-1} X^T \beta$$

$N \quad \downarrow \; p \quad \searrow N$

$N \times p$

$$\begin{pmatrix} y_{1,1,1} \\ \vdots \\ y_{1,1,n_{11}} \\ \vdots \\ y_{a,b,1} \\ \vdots \\ y_{a,b,n_{ab}} \end{pmatrix}$$

$$N = \sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij}$$

These models are not full rank unless we make some restrictions.

usually: i. $\alpha_{i_0} = 0$ for some i_0 or $\sum \alpha_i = 0$,

ii. $\beta_{j_0} = 0$ for some j_0 or $\sum \beta_j = 0$

iii. $\gamma_{i_0 j} = 0$ for one i_0 and all j and $\gamma_{i j_0}$ for one j_0 and all i.

or.

$\sum_{i=1}^{a} \gamma_{i,j} = 0$, for all j and

$\sum_{j=1}^{b} \gamma_{i,j} = 0$, for all i

Estimates under restrictions are unique. for example:
in the balance case with $n_{ij} = n$ with $\geq$ zero

sum constraints are:

$\hat{\mu} = \bar{y}_{...}$ $\quad \hat{\alpha}_i = \bar{y}_{i..} - \bar{y}_{...}$ $\quad \hat{\beta}_i = \bar{y}_{.j.} - \bar{y}_{...}$

and $\hat{\gamma}_{ij} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$

Also consider: $y_{i.j.k} = \mu_{ij} + \varepsilon_{ijk}$

$\hat{\mu}_{ij} = \bar{y}_{ij.}$

Oct. 28.

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

$\begin{cases} H_0: \gamma_{ij} = 0 \text{ for all } i,j \\ H_a: \text{At least } 1\ \gamma_{ij} \neq 0 \end{cases}$

$\begin{cases} H_0: \alpha_i = 0 \text{ for all } i \\ H_a: \text{At least } 1\ \alpha_i \neq 0 \end{cases}$

$\begin{cases} H_0: \beta_j = 0 \text{ for all } j \\ H_a: \text{At least } 1\ \beta_i \neq 0 \end{cases}$

$y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$.

H_0: All μ_{ij}.

Restrictions: $\alpha_1 = 0$, $\beta_1 = 0$

$\gamma_{1j} = 0$, $\gamma_{i1} = 0$.

$\downarrow$ $\quad$ $\downarrow$

for all j $\quad$ for all i.

eg: $\hat{\mu}_{I,B} = \hat{\mu} + \hat{\alpha}_I + \hat{\beta}_B + \hat{\gamma}_{I,B}$.

ANOVA Tables:

- One way ANOVA N: Total # of obs.

Source	DF	SS
Factor A	$a-1$	$N \cdot \bar{y}_{..}^2$ (SS.A)
Error	$N-a$	$\sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i.})^2$ (SSE)

- Two way ANOVA (Balanced Case)

$y_{i,jk} = \mu + \alpha_i + \beta_j + \gamma_{i,j} + \varepsilon_{ijk}$

Source	DF	SS	MS
Factor A	$a-1$	SSA	SSA/(a-1)
Factor B	$b-1$	SSB	SSB/(b-1)
Interaction	$(a-1)(b-1)$	SSI	SSI/(a-1)(b-1)
Error	$ab(n-1)$	SSE	SSE/ab(n-1)

↳ # of each group.

$SSA = bn\sum_i (\bar{y}_{i..} - \bar{y}_{...})^2$

$SSB = an\sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2$

$SSI = n\sum_i\sum_j (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$

$SSE = \sum_i\sum_j\sum_k (y_{ijk} - \bar{y}_{...})^2$

① $y_{ij} = \alpha_i + \beta_i x_{ij} + \varepsilon_{ij}$ $\quad \varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$

① ② $y_{ij} = \alpha + \delta_i + (\beta + \gamma_i) x_{ij} + \varepsilon_{ij}$ $\quad \varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$

① $\hat{\alpha}_1$ = Intercept

② $\hat{\alpha}_2$ = Intercept + Indicator.

Also:

$\hat{\alpha}$ = Intercept $\quad \hat{\delta}_1 = 0 \quad \hat{\delta}_2$ = Indicator.

① : $\hat{\beta}_1$ = Continuous

$\hat{\beta}_2$ = Continuous + Interaction.

Also: $\hat{\beta}$ = Continuous. $\quad \hat{\delta}_1 = 0 \quad \hat{\delta}_2$ = Interaction.

$$\begin{pmatrix} y_{1,1} \\ \vdots \\ y_{1,n_1} \\ y_{2,1} \\ \vdots \\ y_{2,n_2} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & x_{11} \\ \vdots & \vdots & \vdots & \vdots \\ & 1 & 0 & x_{1n_1} \\ & 0 & 1 & x_{21} \\ 1 & \vdots & \vdots & \vdots \\ & 0 & 1 & x_{2n_2} \end{pmatrix} \begin{pmatrix} \alpha \\ \delta_1 \\ \delta_2 \\ \beta \end{pmatrix}$$

PS: If $\delta_1 = 0$, delete 2nd column.

↑

For $y_{ij} = (\alpha + \delta_i) + \beta x_{ij} + \varepsilon_{ij}$

Nov. 4th.

Logistic Regression.

$y_1, \dots, y_n$ binary responses.

$x_1, \dots, x_n$ continuous explanatory variable.

$Pr(Y = y_i \mid X = x_i) = \theta(x_i)$

↳ $\theta(.)$ is a function.

$$\text{logit}(\theta(x_i)) = \log\left(\frac{\theta(x_i)}{1-\theta(x_i)}\right) = \alpha + \beta x_i$$

or: $$\frac{\theta(x_i)}{1-\theta(x_i)} = e^{\alpha + \beta x_i}$$

↳ odds: exponential function of x_i.

Interpretation:

$$\frac{\frac{\theta(x_i+1)}{1-\theta(x_i+1)}}{\frac{\theta(x_i)}{1-\theta(x_i)}} = \frac{e^{\alpha + \beta(x+1)}}{e^{\alpha + \beta x}} = e^{\beta}$$

1 unit increase in x, odds ratio will increase e^{β}.

We can have multiple explanatory variables:

$$\text{Logit}(\theta_i) = \log\left(\frac{\theta_i}{1-\theta_i}\right) = \alpha + \sum_{j=1}^{p} \beta_i x_{i,j}$$

What if we have a binary explanatory variable?

$$e^{\beta} = \frac{\frac{P(Y=1 \mid x=1)}{1-P(Y=1 \mid x=1)}}{\frac{P(Y=1 \mid x=0)}{1-P(Y=1 \mid x=0)}}$$

Once you fit the model, you can look at:

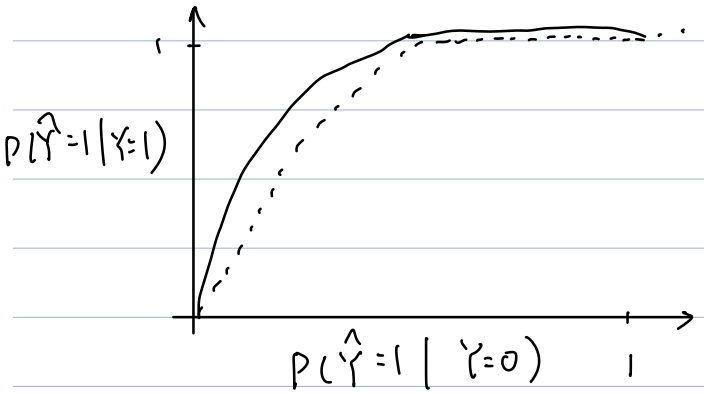
sensitivity:

$$Pr(\hat{Y}=1 \mid Y=1)$$

specificity:

$$Pr(\hat{Y}=0 \mid Y=0)$$

ROC: Receiving Operating characteristic Curve.



Estimation: Maximum Likelihood.

$y_i \overset{ind.}{\sim} \text{Bernoulli}(\theta_i)$

Likelihood:

$$L(\alpha, \beta) = \prod_{i=1}^{n} \theta_i^{y_i} (1-\theta_i)^{1-y_i}$$

$$\text{logit}(\theta_i) = \log\left(\frac{\theta_i}{1-\theta_i}\right) = \alpha + \beta x_i$$

$$\Rightarrow \frac{\theta_i}{1-\theta_i} = e^{\alpha+\beta x_i} \Rightarrow \theta_i = \frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}$$

$$= l(\alpha, \beta) = \sum_{i=1}^{n} y_i \log(\theta_i) + (1-y_i)\log(1-\theta_i)$$

$$= \sum_{i=1}^{n} y_i(\alpha+\beta x_i) + \sum_{i=1}^{n} \frac{1}{e^{\alpha+\beta x_i}}$$