11) $\frac{T}{rm} \overset{\cdot}{\sim} F_{rm, n-mp}$ so use F approx.

ex) $L = (0 \cdots 0 \underset{\uparrow}{1} 0 \cdots 0)$ $\begin{pmatrix} L\underset{\sim}{\beta}_1 \\ \vdots \\ L\underset{\sim}{\beta}_m \end{pmatrix} = \begin{pmatrix} \beta_{1j} \\ \vdots \\ \beta_{mj} \end{pmatrix}$

jth position

$$\begin{pmatrix} L\hat{\underset{\sim}{\beta}}_1 \\ \vdots \\ L\hat{\underset{\sim}{\beta}}_m \end{pmatrix} = \begin{pmatrix} \hat{\beta}_{1j} \\ \vdots \\ \hat{\beta}_{mj} \end{pmatrix} \overset{H_0}{\underset{\sim}{\approx}} N_m(\underset{\sim}{0}, W)$$

$$W_{ik} = \sigma_{ik} (\mathbb{X}^T\mathbb{X})^{-1}_{jj} \quad \text{so } W = (\mathbb{X}^T\mathbb{X})^{-1}_{jj} \underset{\sim}{\Sigma}_{\epsilon}$$

$$\hat{W} = (\mathbb{X}^T\mathbb{X})^{-1}_{jj} \hat{\Sigma}_{\epsilon}.$$

ch 12 not ch 10

12) Suppose you have a univariate DOE ANOVA model $Y_i = x_i^T\beta + e_i$, eg one way ANOVA, two way ANOVA, one way randomized block, $2^k$ factorial, $2^{k-f}$ fractional factorial design. If $m \geq 2$ measurements $Y_{i1}, \ldots, Y_{im}$ (ith case) are taken using the univariate design matrix $\mathbb{X}$, the resulting MANOVA model is the multivariate analog of the univariate DOE ANOVA model.

Ch 10 focuses on the one way ANOVA model for $\mathbb{X}$.

41.5

13] p172-3 randomization: to assign $p$ treatments to $n = n_1 + \cdots + n_p$ units, draw a random permutation of $\{1, \ldots, n\}$. Assign the 1st $n_1$ units to treatment 1, the next $n_2$ to treatment 2, ..., the final $n_p$ units to treatment $p$.

ex] $a_1, a_2, \ldots, a_9$    3 to 3 groups

> sample(9)

6 7 9 / 5 1 4 / 2 8 3

group 1 $a_6, a_7, a_9$

2 $a_5, a_1, a_4$

3 $a_2, a_8, a_3$

see ex 10.1.

14] One way MANOVA test of $H_0\ \underline{\mu_1} = \cdots = \underline{\mu_p}$

Often $\underset{\sim}{\mu_i} = \underset{\sim}{\mu} + \underset{\sim}{\tau_i}$. So $H_0: \underset{\sim}{\tau_1} = \cdots = \underset{\sim}{\tau_p}$.

Assume software gives test statistic to and pval. 4 step test i) $H_0: \underset{\sim}{\mu_1} = \cdots = \underset{\sim}{\mu_p}$ $H_A$ not $H_0$

ii) to

iii) pval

iv) If pval $\le \alpha$, reject $H_0$, conclude not all treatment means are equal.

If pval $> \alpha$, fail to reject $H_0$, conclude all treatment means are equal (or that there is not enough evidence to conclude that not all treatment means are equal). Give a nontechnical sentence if possible.

ex) ( p240-2) Nursing homes are private, nonprofit or government (3 groups). Let $x_1$ = cost of nursing labor $x_2$ = cost of dietary labor, $x_3$ = cost of plant operation and maintenance labor, $x_4$ = cost of housekeeping and laundry labor. and suppose $t_0 = 17.67$, $n = 510$, $p = 4$, and pval $= P(17.67 < F_{8,1020}) = 0.000$. Do a 4 step oneway MANOVA test.

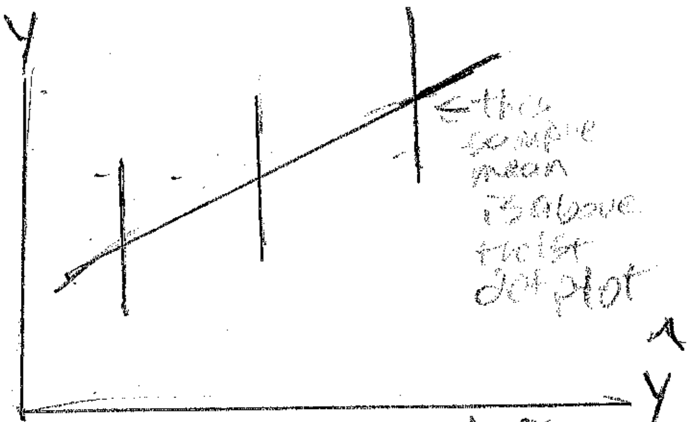
soln i) Ho $\underline{\mu}_1 = \cdots = \underline{\mu}_3$ $H_1$ not Ho

ii) $t_0 = 17.67$

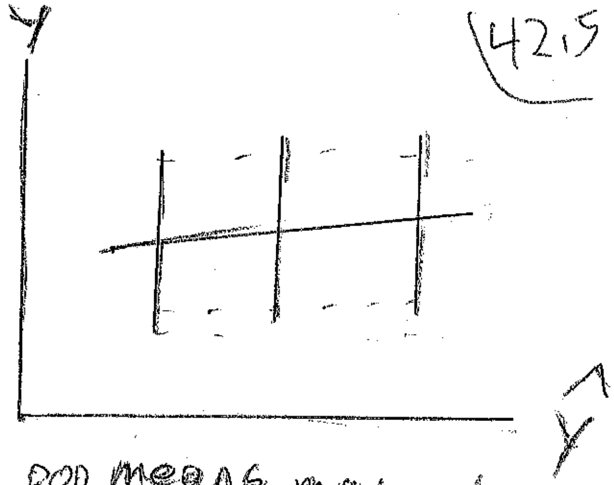
iii) pval $= 0.000$

iv) reject Ho the mean costs for the 3 types of nursing homes are different.

15)




pop means are different




pop means may not be different
inconclusive



# Ch 11 Factor Analysis

1) Factor analysis

i) $\hat{\Sigma} \approx \hat{L}\hat{L}^T + \hat{\psi}$

so $\hat{\Sigma} \approx \hat{L}\hat{L}^T$ if $\hat{\psi}$ is small,

ii) cluster variables into groups called factors,

iii) suggests that the groups are factors that explain the dispersion more simply than $x_1, \ldots, x_p$.

2) Given $R$ (sample corr matrix), put variables into groups.
Start with the 2 variables with the highest $r_{ij}$ provided $|r_{ij}| \geq c$. Then add the variable with the highest $|r_{ij}| \geq c$ with variables in the group. Use $c = 0.9, 0.8, 0.7, 0.6, 0.5$ etc.
If all $|r_{ij}| < c$, then each group has 1 variable.

ex]

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 classics | | .83 | .78 | .7 | .66 | .63 |
| 2 French | | | .67 | .67 | .65 | .57 |
| 3 English | | | | .64 | .54 | .51 |
| 4 math | | | | | .45 | .51 |
| 5 discrimination | | | | | | .4 |
| 6 music | | | | | | 1 |

$\hat{R}$ = sample corr matrix

If $0 < c \le 0.4$, put all variables in 1 group.
If $c = 0.83$, classics, French = group 1
If $0.7 < c \le 0.78$, group 1 = classics, French, English = "language factor"
2 = math
3 = discrimination
4 = music

If $0.67 \le c < 0.7$, group 1 = classics, French, English, math

If $c = 0.66$ group 1 = 1–5
group 2 = 6

3] Factor analysis output is a lot like PCA output, but replace PC1, PC2, ..., PCP by Factor 1, Factor 2, ..., Factor k.
promax rotation tries to give loading 0 to many variables. variables with nonzero loadings are important.

Want cumulative variance $> c$ where $c = 0.5, 0.6, \ldots, 0.9$ are typical.

4) There are several methods of factor analysis. R uses MLE factor analysis (with several types of rotations). This method can use data or a dispersion matrix $S$ or $R$. As a diagnostic, use the RMVN dispersion matrix or generalized correlation matrix.

< Tend to get this output when R wants more factors.

5] (If the 1 factor model is not sufficient) R will give a test for whether a k factor model is sufficient. A k factor model with pval $< 0.05$ is not sufficient; more factors are needed. A k factor model with pval $> 0.05$ is sufficient.

not sure: for some data sets pval $< .05$ seems good) and for some pval $> .05$ seems good. See E3 rev (14).

$H_0$: k factors are sufficient $H_1$: not $H_0$

See factor Analysis handout

6] A k factor model "makes sense" if the degrees of freedom $d \geq 0$ where $d = \frac{1}{2}(p-k)^2 - \frac{1}{2}(p+k)$.

$d < 0$ doesn't make sense

ex) $p=5$,

| $k$ | 1 | 2 | 3 |
|---|---|---|---|
| $d$ | 5 | 1 | -2 |

$k=1$: $\frac{1}{2}(5-1)^2 - \frac{1}{2}(5+1) = 8-3=5$

$k=2$: $\frac{1}{2}(5-2)^2 - \frac{1}{2}(5+2) = \frac{9}{2} - \frac{7}{2} = 1$

$k=3$: $\frac{1}{2}(5-3)^2 - \frac{1}{2}(5+3) = 2-4=-2$

So if the factor analysis model is appropriate for a set of 5 variables, it will have no more than 2 factors.

7) Factor analysis is a "wow that makes sense" method.

PCA, MLE

8) There are at least 2 methods of factor analysis that can be robustified by applying the factor analysis method on the RMVN subset.

9) $\underset{p\times1}{\underset{\sim}{X}} = \underset{p\times1}{\underset{\sim}{\mu}} + \underset{p\times m}{L} \underset{m\times1}{\underset{\sim}{F}} + \underset{p\times1}{\underset{\sim}{\varepsilon}}$, $\Sigma_{\underset{\sim}{X}} \approx LL^T + \Psi = \Sigma_{\underset{\sim}{F}}$

where there is equality for diagonal elements.

$\underset{\sim}{F} = (F_1, \ldots, F_m)^T$ is a vector of unknown common factors.

$\underset{\sim}{\varepsilon}$ is the vector of unique factors and $1 \le m < p$.

$E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$

$\Psi = \mathrm{Cov}(\underset{\sim}{\varepsilon}) =$ error covariance matrix

$E(\underset{\sim}{F}) = \underset{\sim}{0}, \ \mathrm{Cov}(\underset{\sim}{F}) = I_m.$

The diagonal elements of $LL^T = h_i^2$

$= \sum_{j=1}^{k} L_{ij}^2$ are called communalities.

The diagonal elements of $\psi$ are $\psi_{ii}$

and are called uniquenesses.

10] Let $\hat{\Gamma}$ be an orthogonal matrix. Then

$\hat{L}^* (\hat{L}^*)^T = (\hat{L}\hat{\Gamma})(\hat{L}\hat{\Gamma})^T = \hat{L}\hat{\Gamma}\hat{\Gamma}^T\hat{L}^T = \hat{L}\hat{L}^T.$

varimax and promax rotations find $\hat{\Gamma}$

so that $\hat{L}^* = \hat{L}\hat{\Gamma}$ has loadings that

are easier to interpret than the loadings of $\hat{L}$.

11] PCA factor analysis:

$\hat{\Sigma} = \sum_{i=1}^{p} \hat{\lambda}_i \hat{\underset{\sim}{e}}_i \hat{\underset{\sim}{e}}_i^T$. Let $\hat{L} = \left[\sqrt{\hat{\lambda}_1}\, \hat{\underset{\sim}{e}}_1, \ldots, \sqrt{\hat{\lambda}_m}\, \hat{\underset{\sim}{e}}_m\right]$.

Take $\hat{\Sigma} \approx \hat{L}\hat{L}^T + \hat{\Psi} = \hat{\Sigma}_{\underset{\sim}{F}}$ where $\hat{L}\hat{L}^T = \sum_{i=1}^{m} \hat{\lambda}_i \hat{\underset{\sim}{e}}_i \hat{\underset{\sim}{e}}_i^T$

and $\hat{\Psi} = \mathrm{diag}(\hat{\Psi}_1, \ldots, \hat{\Psi}_p)$ where $\hat{\Psi}_i = \hat{\Sigma}_{ii} - (\hat{L}\hat{L}^T)_{ii}$.

$\hat{L}$ is the matrix of factor loadings.

So Factor k $= \sqrt{\hat{\lambda}_k}\, \hat{e}_k$ for PCA factor analysis.

The ith estimated communality

$$\hat{h}_i^2 = \hat{l}_{i1}^2 + \hat{l}_{i2}^2 + \cdots + \hat{l}_{im}^2$$

$$\hat{\sigma}_{ii} = \hat{h}_i^2 + \hat{\psi}_i \quad \text{for } i=1,\ldots,p.$$

12) $\hat{\Sigma} - \hat{\Sigma}_F$ is useful.

DD plot of $D_i(\overline{w}, \hat{\Sigma})$ vs $D_i(\overline{w}, \hat{\Sigma}_F)$ may be useful,

where $\overline{w} = \begin{cases} \overline{x} & \text{if } \hat{\Sigma} = S \\ \overline{z} & \text{if } \hat{\Sigma} = R \end{cases}$

$z_i$ = standardized predictors

Ch 12 1) The multivariate linear regression (mreg) model is a special case of the multivariate linear model where at least one predictor $x_j$ is continuous. The MANOVA model is a multivariate linear model where all of the predictors are categorical variables, often coded as indicator variables.

2) The mreg model is $y_i = B^T x_i + \epsilon_i$ for $i = 1,\ldots,n$ with $m \geq 2$ response variables $Y_1,\ldots,Y_m$ and $p$

predictor variables $x_1 \equiv 1, \underbrace{x_2, x_3, \ldots, x_p}_{\text{non trivial predictors}}$. 45.5

ith case $= (\underset{\sim}{x}_i^T, \underset{\sim}{y}_i^T)$ but $x_{i1} = 1$ is often omitted

data matrix $\begin{pmatrix} \underset{\sim}{x}_1^T, \underset{\sim}{y}_1^T \\ \vdots \\ \underset{\sim}{x}_n^T, \underset{\sim}{y}_n^T \end{pmatrix}$ but column of 1's is usually omitted.

Let $B = [\underset{\sim}{\beta}_1 \ \underset{\sim}{\beta}_2 \cdots \underset{\sim}{\beta}_m]$, $E = \begin{pmatrix} \underset{\sim}{\varepsilon}_1^T \\ \vdots \\ \underset{\sim}{\varepsilon}_n^T \end{pmatrix} = [\underset{\sim}{e}_1 \cdots \underset{\sim}{e}_m]$, $\underset{n \times m}{Z} = [\underset{\sim}{Y}_1 \ \underset{\sim}{Y}_2 \cdots \underset{\sim}{Y}_m]$.

In matrix form, $\underset{n \times m}{Z} = \underset{n \times p}{X} \underset{p \times m}{B} + \underset{n \times m}{E}$

$E(\underset{\sim}{\varepsilon}_k) = \underset{\sim}{0}$, $Cov(\underset{\sim}{\varepsilon}_k) = \Sigma_{\varepsilon} = (\sigma_{ij})$, $k = 1, \ldots, n$

$E(\underset{\sim}{e}_i) = \underset{\sim}{0}$, $Cov(\underset{\sim}{e}_i, \underset{\sim}{e}_j) = \sigma_{ij} I_n$, for $i, j = 1, \ldots, m$.

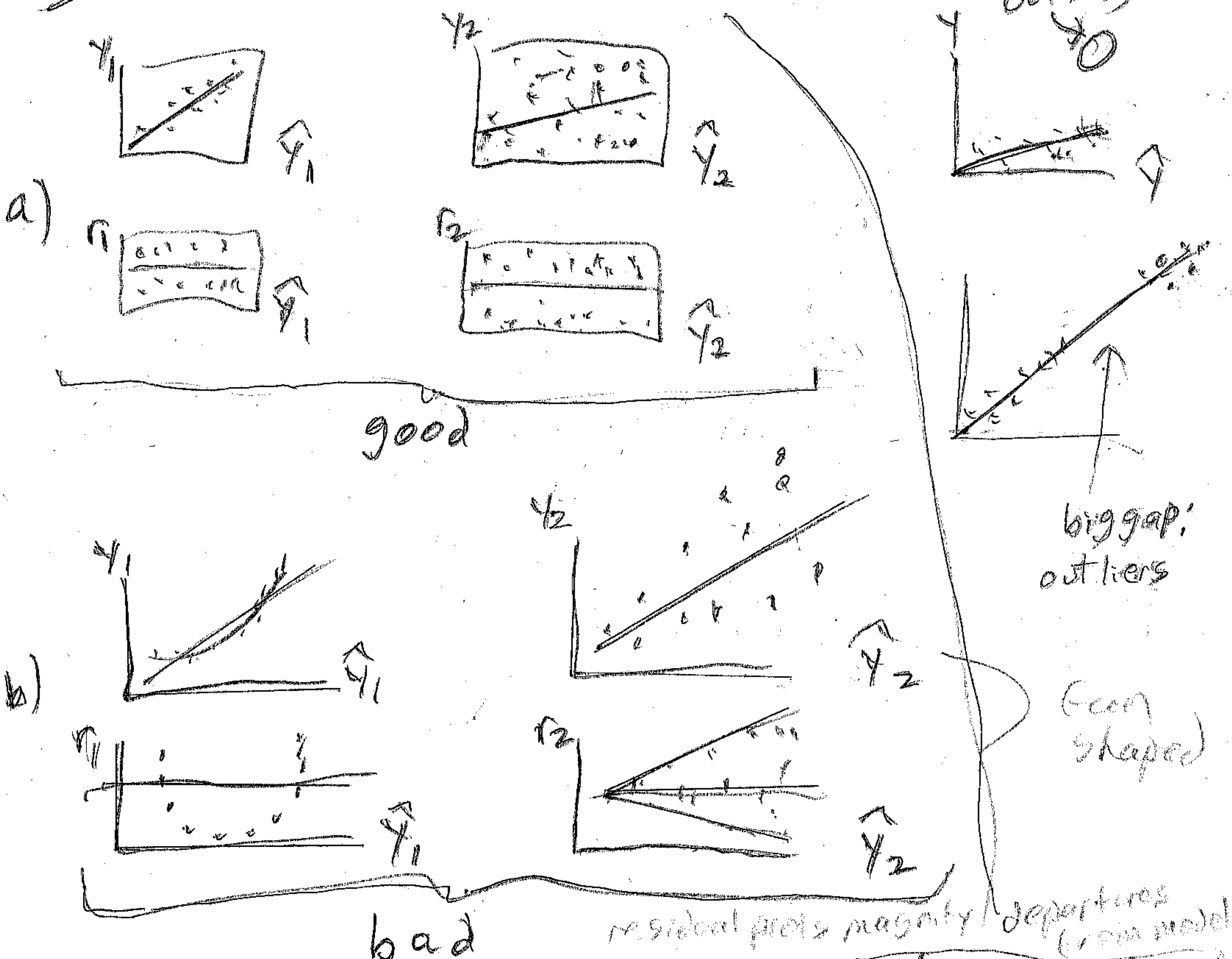
3) Each response variable in the mreg model follows a univariate multiple linear regression (MLR) model $\underset{n \times 1}{\underset{\sim}{Y}_j} = \underset{n \times p}{X} \underset{p \times 1}{\underset{\sim}{\beta}_j} + \underset{n \times 1}{\underset{\sim}{e}_j}$ $\quad j = 1, \ldots, m$

same $X$ for each $j$

4) For each variable $Y_k$, make a response plot of $\hat{Y}_{ik}$ versus $Y_{ik}$ and a residual plot of $\hat{Y}_{ik}$ vs $r_{ik} = Y_{ik} - \hat{Y}_{ik}$. If $n > 10p$ and the mreg model is appropriate, the plotted points in each response plot should scatter in a roughly evenly populated band about the identity line. The plotted points in the residual plot should scatter

in a roughly evenly populated band about the $r = 0$ line.

5] Curvature or fan shaped plots are bad.




residual plots magnify departures from model

6) Make a scatter plot matrix of $Y_1, \ldots, Y_m$ and the continuous predictors. Use power transformations, ladder rule, and log rule to remove strong nonlinearities.

46.5

7) Test $H_0\ LB=0$ vs $H_1\ LB\neq 0$

$\underset{r\times p}{L}$ $r\leq p$ full rank $r$

$W_e = \hat{E}^T E = (n-p)\hat{\Sigma}_{\epsilon}$

$H = \hat{B}^T L^T \left[L(X^T X)^{-1} L^T\right]^{-1} L\hat{B}$

$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$ the eigenvalues of $W_e^{-1} H$.

Wilks' $\Lambda(L) = |(H+W_e)^{-1} W_e| = \prod_{i=1} (1+\lambda_i)^{-1}$

Pillai's $V(L) = tr\left((H+W_e)^{-1} H\right) = \sum_{i=1}^{m} \frac{\lambda_i}{1+\lambda_i}$

Hotelling Lawley $U(L) = tr(W_e^{-1} H) = \sum_{i=1}^{m} \lambda_i$

$= \frac{1}{n-p}\left[vec(L\hat{B})\right]^T \left[\hat{\Sigma}_{\epsilon}^{-1} \otimes \left(L(X^T X)^{-1} L^T\right)^{-1}\right] vec(L\hat{B})$

Roy's $\lambda_{max}(L) = \lambda_1(L)$

8) Under regularity conditions

$-\left[n-p+1-0.5(m-r+3)\right]\log \Lambda(L) \xrightarrow{D} \chi^2_{rm}$

$(n-p)\, V(L) \xrightarrow{D} \chi^2_{rm}$

$(n-p)\, U(L) \xrightarrow{D} \chi^2_{rm}$

if $h = \max(r,m)$, $\frac{n-p-h+r}{h}\lambda_{max}(L) \underset{\uparrow \text{ bad if } r>1}{\approx} F(h, n-p-h-r)$