Stat 306:
Finding Relationships in Data.

Lecture 23

6.2 Count Regression

a.k.a Non-negative integers!

Counts!

0, 1, 2, 3, 4, ...

9, 10, 11, 12, ...

34, 35, 36 ....

Count Regression

Y, the outcome variable, is a count,
i.e. a non-negative integer.

X is any explanatory covariate.

Count Regression

$$Y_i \sim \text{Poisson}(\lambda) \text{ with mean } \lambda > 0$$

We want to model the mean, λ , as a function of covariates X:

$$Y_i \sim \text{Poisson}(\lambda(\mathbf{x}_i)), \text{ where } \lambda(\mathbf{x}_i) = \exp\{\mathbf{x}_i^T \boldsymbol{\beta}\} \text{ or}$$

$$\log \lambda(\mathbf{x}_i) = \mathbf{x}_i^T \boldsymbol{\beta}.$$

Important property of the Poisson distribution:

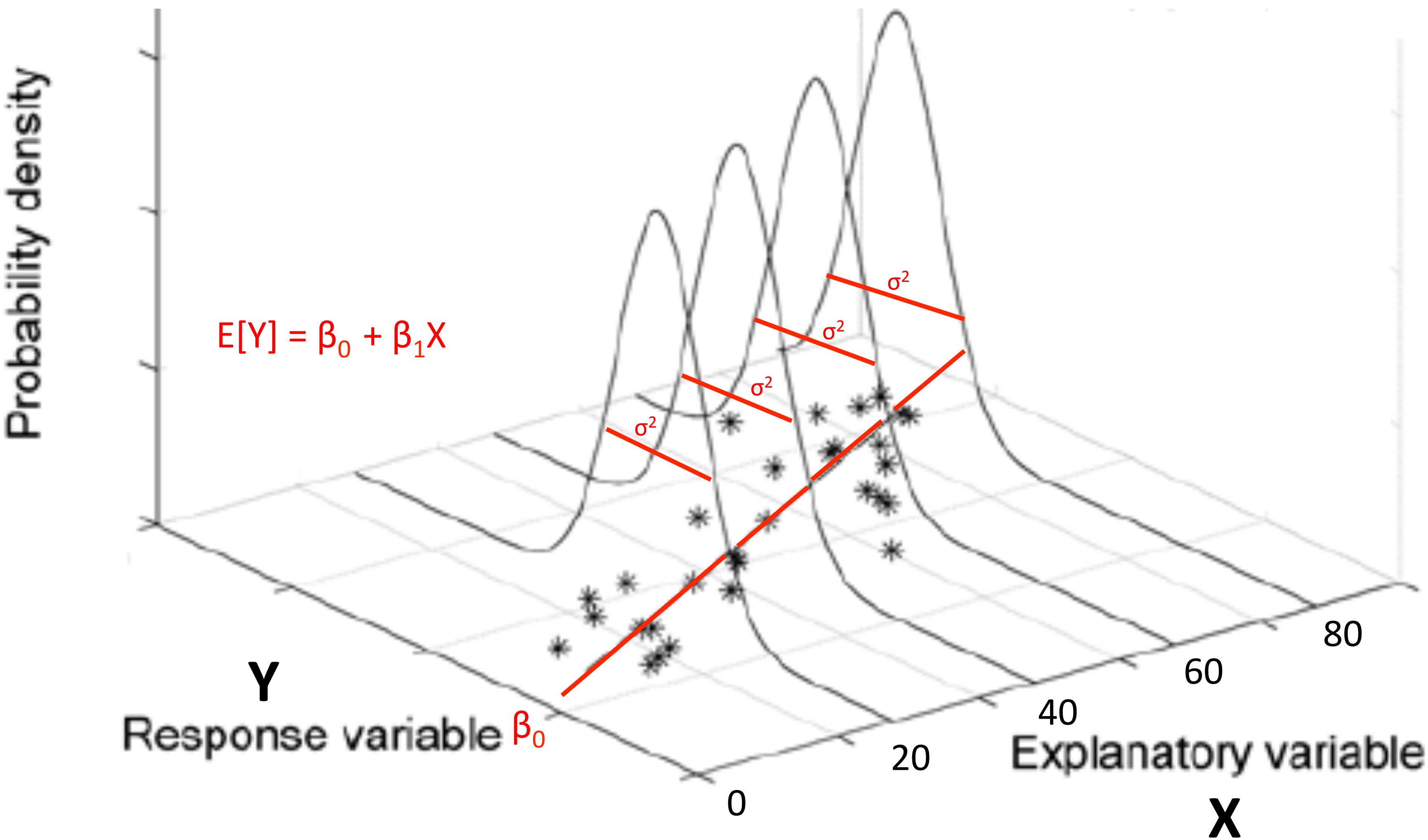
E[Y] = λ
Var[Y] = λ

Important property of the Poisson distribution:

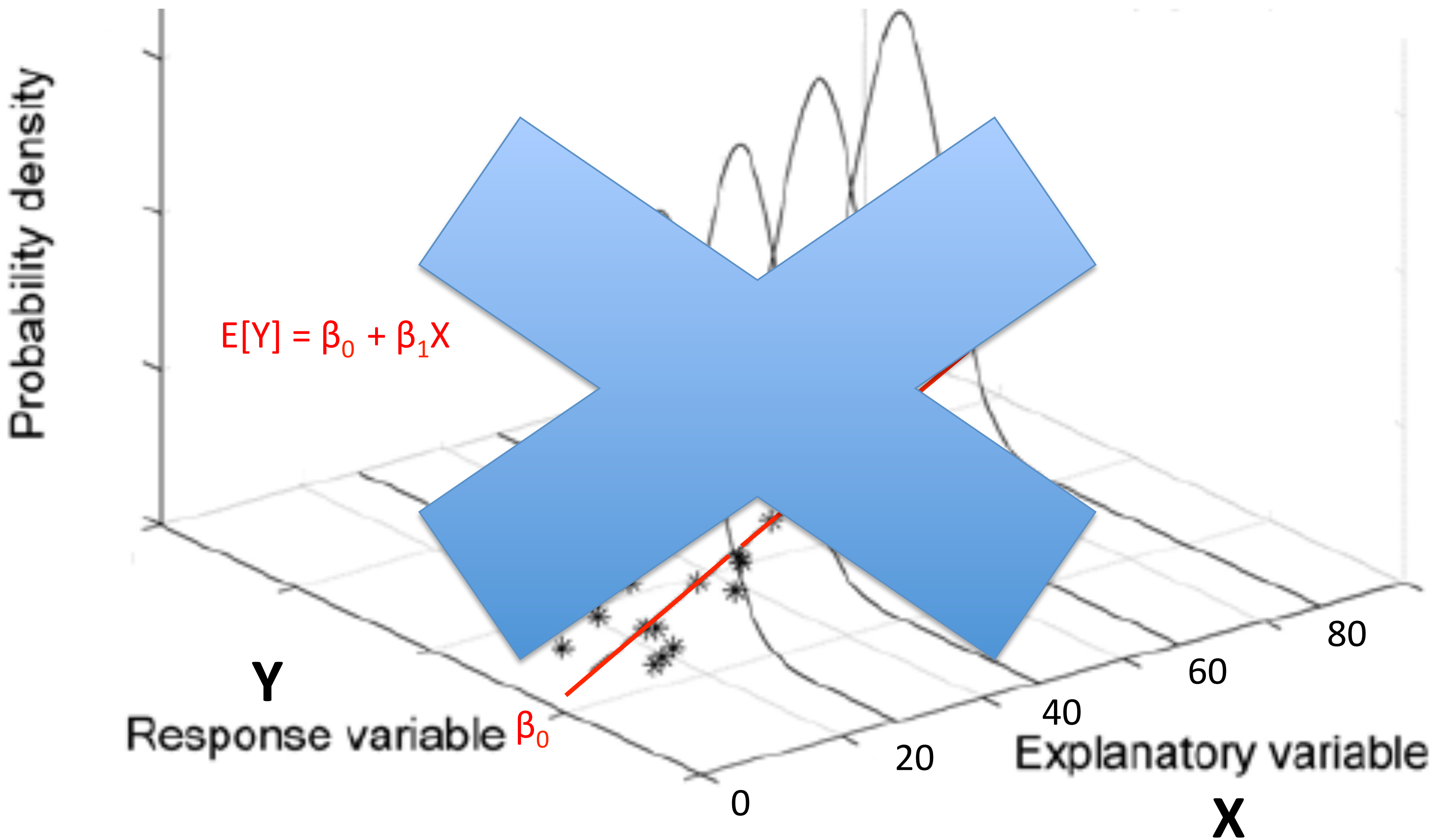
$$E[Y] = \lambda$$
$$Var[Y] = \lambda$$

Homoscedasticity?

Section 2.2 - Statistical linear regression model



Section 2.2 - Statistical linear regression model



Examples of general regression

1. Usual regression (normal or Gaussian response): $Y_i \sim N(\mu, \sigma^2)$ is extended to $Y \sim N(\mathbf{x}_i^T\boldsymbol{\beta}, \sigma^2)$. That is, $\mu(\mathbf{x}_i) = \mu_i = \mathbf{x}_i^T\boldsymbol{\beta}$.

2. Binary regression (logistic regression as special case): $Y_i \sim$Bernoulli(π), where $\pi = \Pr(Y_i = 1)$ and $1 - \pi = \Pr(Y_i = 0)$, is extended to $Y_i \sim$ Bernoulli$(\pi(\mathbf{x}_i))$, where $\pi(\mathbf{x}_i) = \exp\{\mathbf{x}_i^T\boldsymbol{\beta}\}/[1 + \exp\{\mathbf{x}_i^T\boldsymbol{\beta}\}$ or

$$\log\left\{\frac{\pi(\mathbf{x}_i)}{1 - \pi(\mathbf{x}_i)}\right\} = \mathbf{x}_i^T\boldsymbol{\beta}.$$

   [Logistic cdf is $F(z) = e^z/(1 + e^z)$, $-\infty < z < \infty$]

3. Count regression (Poisson regression as special case): $Y_i \sim$ Poisson(λ) with mean $\lambda > 0$ and $P(Y_i = y) = \lambda^y e^{-\lambda}/y!$ $(y = 0, 1, \ldots)$ is extended to $Y_i \sim$ Poisson$(\lambda(\mathbf{x}_i))$, where $\lambda(\mathbf{x}_i) = \exp\{\mathbf{x}_i^T\boldsymbol{\beta}\}$ or

$$\log \lambda(\mathbf{x}_i) = \mathbf{x}_i^T\boldsymbol{\beta}.$$

   Poisson regression can be used for insurance claim data to model the number of car accidents (or claims) per year for individuals as a function or demographic and risk factors.

Maximum Likelihood Estimation

Steps to get Maximum Likelihood Estimates:

1. Define the Likelihood as function of the parameters given the data.
2. Define the log-Likelihood.
3. Maximize the log-Likelihood or minimize the negative log-Likelihood.

For Poisson regression, we have:

$$L(\boldsymbol{\beta}; data) = \prod_{i=1}^{n} [\lambda(\mathbf{x}_i)]^{y_i} \frac{e^{-\lambda(\mathbf{x}_i)}}{y_i!}$$

and:

$$\begin{aligned} \log L(\boldsymbol{\beta}; data) &= \sum_{i=1}^{n} \{y_i \log \lambda(\mathbf{x}_i) - \lambda(\mathbf{x}_i) - \log(y_i!)\} \\ &= \sum_{i=1}^{n} \{y_i \mathbf{x}_i^T \boldsymbol{\beta} - \exp(\mathbf{x}_i^T \boldsymbol{\beta}) - \log(y_i!)\} \end{aligned}$$

Unfortunately…

There is no closed form solution, but statistical software obtain $\hat{\beta}_0, \hat{\beta}_1$ with an iterative method.

Maximum Likelihood Estimation

Poisson: $P(Y=y;\mathbf{x}) = [\lambda(\mathbf{x})]^y e^{-\lambda(\mathbf{x})}/y!$, $\lambda(\mathbf{x}) = \exp\{\mathbf{x}^T\boldsymbol{\beta}\}$ (intercept 1 included in $\mathbf{x}$).
Data $(\mathbf{x}_i, y_i)$, $i=1,\dots,n$; $y_i \in \{0,1,2,\dots\}$.
Poisson likelihood in $\boldsymbol{\beta}$ is:

$$L(\boldsymbol{\beta}; data) = \prod_{i=1}^{n} [\lambda(\mathbf{x}_i)]^{y_i} \frac{e^{-\lambda(\mathbf{x}_i)}}{y_i!}$$

Loglikelihood in $\boldsymbol{\beta}$ is:

$$\begin{aligned}\log L(\boldsymbol{\beta}; data) &= \sum_{i=1}^{n}\{y_i \log\lambda(\mathbf{x}_i) - \lambda(\mathbf{x}_i) - \log(y_i!)\} \\ &= \sum_{i=1}^{n}\{y_i \mathbf{x}_i^T\boldsymbol{\beta} - \exp(\mathbf{x}_i^T\boldsymbol{\beta}) - \log(y_i!)\}\end{aligned}$$

For the null model of no effect for explanatory variables, $\lambda = e^{\beta_0}$ for all i, log-likelihood is:

$$\sum_{i=1}^{n}\{y_i\log\lambda - \lambda - \log(y_i!)\} = y_+ \log\lambda - n\lambda - \sum_{i=1}^{n}\log(y_i!),$$

with maximum likelihood estimate (MLE) $\hat{\lambda} = \overline{y}$.

Maximum Likelihood Estimation

Steps to get standard error for the Maximum Likelihood Estimates (MLE):

1. Take the second derivative of the log-Likelihood function. This is the Hessian Matrix. The negative of the Hessian is the Fisher information matrix.

2. Evaluate the Fisher Information matrix at the MLE. This is known as the “observed Fisher Information matrix”.

3. Take the inverse of the “observed Fisher Information matrix”. This is your estimate for the Variance-Covariance matrix of your parameter. The diagonal elements are the estimated Variances.

4. Take the square root of the diagonal elements to obtain the standard errors.

Maximum Likelihood Estimation

Gradient and Hessian.

$$\log L(\boldsymbol{\beta}; data) = \sum_{i=1}^{n} \{y_i \mathbf{x}_i^T \boldsymbol{\beta} - \exp(\mathbf{x}_i^T \boldsymbol{\beta}) - \log(y_i!)\}$$

$$\frac{\partial \log L(\boldsymbol{\beta}; data)}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \{\mathbf{x}_i y_i - \mathbf{x}_i \exp(\mathbf{x}_i^T \boldsymbol{\beta})\}$$

$$\frac{-\partial^2 \log L(\boldsymbol{\beta}; data)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T \exp(\mathbf{x}_i^T \boldsymbol{\beta})$$

SEs for $\hat{\beta}$s come from the square root of the diagonal elements of

$$\left[\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T \exp(\mathbf{x}_i^T \boldsymbol{\beta}) \,\Big|_{\hat{\boldsymbol{\beta}}} \right]^{-1}$$

Diagnostics for Poisson regression: **The Deviance.**

The residual deviance becomes ($\sum_{i=1}^n \log(y_i!)$ cancels and $\hat{\lambda}_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$)

$$2[\ell(\lambda_1^{(S)}, \ldots, \lambda_n^{(S)}) - \ell(\hat{\lambda}_1, \ldots, \hat{\lambda}_n)] = 2\left[\sum_{i=1}^n \{y_i \log y_i - y_i\} - \sum_{i=1}^n \{y_i \mathbf{x}_i^T \boldsymbol{\beta} - \exp(\mathbf{x}_i^T \boldsymbol{\beta})\}\right].$$

The null deviance becomes ($\hat{\lambda}_i = \overline{y}$)

$$2[\ell(\lambda_1^{(S)}, \ldots, \lambda_n^{(S)}) - \ell(\overline{y}, \ldots, \overline{y})] = 2\left[\sum_{i=1}^n \{y_i \log y_i - y_i\} - \{n\overline{y} \log \overline{y} - n\overline{y}\}\right] = 2\left[\sum_{i=1}^n y_i \log y_i - n\overline{y} \log \overline{y}\right].$$

Here $0 \log 0 = 0$ as limit of $z \log z$ as $z \to 0^+$.

$$0 \leq \text{residual deviance} \leq \text{null deviance because } \ell(\text{saturated}) \geq \ell(\text{explanatory}) \geq \ell(\text{no explanatory}).$$

Diagnostics for Poisson regression: **The AIC**

1. The AIC seeks to identify "good models" by considering both the likelihood and the number of parameters in the model:

$$AIC = -2 \cdot log(Likelihood) + 2 \cdot (\# \text{ of parameters})$$

2. As such it is similar to the adjusted R^2.

3. Lower **AIC** suggests that the model is better.

Interpretation for Poisson regression:

Suppose model is $P(Y_i = y; \mathbf{x}_i) = e^{-\lambda_i}\lambda_i^y/y!$, where λ_i depends on $\mathbf{x}_i$. The model-based expected number of occurrences of the k claims is

$$E_k = \sum_{i=1}^{n} \mathrm{E}\left[I(Y_i = k)\right] = \sum_{i=1}^{n} P(Y_i = k; \mathbf{x}_i) = \sum_{i=1}^{n} e^{-\lambda_i}\lambda_i^k/k!.$$

For a fitted model, replace λ_i by $\hat{\lambda}_i = \exp\{\mathbf{x}_i^T\hat{\boldsymbol{\beta}}\}$ from the regression model.

Expected number of 0s from the model is $\sum_{i=1}^{n} e^{-\hat{\lambda}_i}$.

Expected number of 1s from the model is $\sum_{i=1}^{n} e^{-\hat{\lambda}_i}\hat{\lambda}_i$.

Expected number of 2s from the model is $\sum_{i=1}^{n} e^{-\hat{\lambda}_i}\hat{\lambda}_i^2/2$.

Excellent information on Poisson Regression:

https://freakonometrics.hypotheses.org/9593

https://freakonometrics.hypotheses.org/2289

The Table for Generalized Linear Regression: Logistic Regression

Step 0:
For θ, define
the log-likelihood.

Step 1:
Maximize the log-likelihood
to find, MLE: $\hat{\theta}$

Step 2:
Determine the negative
Hessian matrix

Step 3:
Define
se($\hat{\theta}$) =
estimate of $\sqrt{\mathrm{Var}(\hat{\Theta})}$

Step 4:
Define
(1-α)% C.I. =
$\hat{\theta} \pm c \times se(\hat{\theta})$

Population parameter or **"something we would like to estimate"**	**Log-likelihood**	**MLE**	**The negative Hessian**	**Variance of the estimator**	**Standard Error of estimator**	**Confidence Interval**
β	$\frac{\partial \log L(\boldsymbol{\beta}; \mathrm{data})}{\partial \boldsymbol{\beta}}$	**Numerical methods**	**H** $= \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T \pi_i (1-\pi_i)$	Var[**B**] = $(\mathrm{H})^{-1}$	se(**b**) = $diag(\sqrt{(H^{-1}(\hat{\beta}))})$	C.I. for **β**
$\pi(x)$	$\hat{\pi}(x_i) = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \ldots + \hat{\beta}_p x_{ip}}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \ldots + \hat{\beta}_p x_{ip}}}$		Delta method? http://www.indiana.edu/~jslsoc/stata/ci_computations/spost_deltaci.pdf			

The Table for Generalized Linear Regression: Poisson Regression

Step 0:
For θ, define
the log-likelihood.

Step 1:
Maximize the log-likelihood
to find, MLE: $\hat{\theta}$

Step 2:
Determine the negative Hessian matrix

Step 3:
Define
se($\hat{\theta}$) =
estimate of $\sqrt{\text{Var}(\hat{\Theta})}$

Step 4:
Define
(1-α)% C.I. =
$\hat{\theta} \pm c \times se(\hat{\theta})$

Population parameter or "something we would like to estimate"	**Log-likelihood**	**MLE**	**The negative Hessian**	**Variance of the estimator**	**Standard Error of estimator**	**Confidence Interval**
β	$\frac{\partial \log L(\boldsymbol{\beta}; \text{data})}{\partial \boldsymbol{\beta}}$	**Numerical methods**	**H** $= \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T \exp(\mathbf{x}_i^T \boldsymbol{\beta})$	Var[**B**] = $(H)^{-1}$	se(**b**) = $diag(\sqrt{(H^{-1}(\hat{\beta}))})$	C.I. for **β**
$\lambda(\mathbf{x}_i) = \exp\{\mathbf{x}_i^T \boldsymbol{\beta}\}$			Delta method? http://www.indiana.edu/~jslsoc/stata/ci_computations/spost_deltaci.pdf			