2 Analisis con graficos

Caso 1

Fisher and Belle (1993) report mortality rates due to malignant melanoma of the skin for white males during the period 1950–1969, for each state on the US mainland. The data are given in Table 2.1 and include the number of deaths due to malignant melanoma in the corresponding state, the longitude and latitude of the geographic centre of each state, and a binary variable indicating contiguity to an ocean, that is, if the state borders one of the oceans. Questions of interest about these data include: how do the mortality rates compare for ocean and non-ocean states? and how are mortality rates affected by latitude and longitude?

View(Usmelanoma)

- Normalizar el eje x

```
> xr <- range(USmelanoma$mortality) * c(0.9, 1.1)
> xr
```

- Histograma y boxplot

```
> layout(matrix(1:2, nrow = 2))
> par(mar = par("mar") * c(0.8, 1, 1, 1))
> boxplot(USmelanoma$mortality, ylim = xr, horizontal = TRUE,
+ xlab = "Mortality")
> hist(USmelanoma$mortality, xlim = xr, xlab = "", main = "",
+ axes = FALSE, ylab = "")
> axis(1)
```

- Boxplot por tipo de estado

```
boxplot(mortality ~ ocean, data = USmelanoma,
+ xlab = "Contiguity to an ocean", ylab = "Mortality")
```

- Densidades condicionales estimadas

```
> dyes <- with(USmelanoma, density(mortality[ocean == "yes"]))
> dno <- with(USmelanoma, density(mortality[ocean == "no"]))
> plot(dyes, lty = 1, xlim = xr, main = "", ylim = c(0, 0.018))
> lines(dno, lty = 2)
> legend("topleft", lty = 1:2, legend = c("Coastal State",
+ "Land State"), bty = "n")
```

- Distribucion geografica, scatterplots

```
> layout(matrix(1:2, ncol = 2))
> plot(mortality ~ longitude, data = USmelanoma)
> plot(mortality ~ latitude, data = USmelanoma)
```

- Discriminado estados

```
> plot(mortality ~ latitude, data = Usmelanoma, pch = as.integer(USmelanoma
$ocean))
> legend("topright", legend = c("Land state", "Coast state"),
pch = 1:2, bty = "n")
```

Caso 2

sample. The Chinese Health and Family Life Survey sampled 60 villages and urban neighbourhoods chosen in such a way as to represent the full geographical and socioeconomic range of contemporary China excluding Hong Kong and Tibet. Eighty-three individuals were chosen at random for each location from official registers of adults aged between 20 and 64 years to target a sample of 5000 individuals in total. Here, we restrict our attention to women with current male partners for whom no information was missing, leading to a sample of 1534 women with the following variables (see Table 2.2 for example data sets):

R_edu: level of education of the responding woman,

R_income: monthly income (in yuan) of the responding woman,

R_health: health status of the responding woman in the last year,

R_happy: how happy was the responding woman in the last year,

A_edu: level of education of the woman's partner,

A_income: monthly income (in yuan) of the woman's partner.

In the list above the income variables are continuous and the remaining variables are categorical with ordered categories. The income variables are based on (partially) imputed measures. All information, including the partner's income, are derived from a questionnaire answered by the responding woman only. Here, we focus on graphical displays for inspecting the relationship of these health and socioeconomic variables of heterosexual women and their partners.

- Frecuencias de variables categoricas

```
barplot(xtabs(~ R_happy, data = CHFLS))
```

- Spineplot: categoricas vs categoricas

```
xtabs(~ R_happy + R_health, data = CHFLS)

plot(R_happy ~ R_health, data = CHFLS)
```

- Spinegram: categorias vs. numericas

```
> layout(matrix(1:2, ncol = 2))
> plot(R_happy ~ log(R_income + 1), data = CHFLS)
> cdplot(R_happy ~ log(R_income + 1), data = CHFLS)
```

- Relaciones entre los salarios de la pareja

```
xyplot(jitter(log(A_income + 0.5)) ~ jitter(log(R_income + 0.5)) |
R_edu, data = CHFLS)
```