

MLE COURSE: Homework #2.

1. Which of these is the correct interpretation of a $(1 - \alpha)$ confidence interval?

- An interval that has a $1 - \alpha\%$ chance of containing the true value of the parameter.
- An interval that over $1 - \alpha\%$ of replications contains the true value of the parameter, *on average*.

What interpretation do people really *want*.

2. Explain what the following S language function performs:

```
vector.test <- function(in.vector, base.vector) {  
  accept <- 0  
  for(i in 1:length(in.vector)) {  
    if ( (in.vector[i] > 0) && (in.vector[i] < base.vector[i]) )  
      accept <- 1  
  }  
  return(accept)  
}
```

3. The first column in the dataset `gov.employ-1998.dat` <http://artsci.wustl.edu/~jgill/data/gov-employ-1998.dat>) counts percent blacks in the federal workforce by agency. Change this variable into a factor with the categories low=0-15, medium=15.1-30, and high=31+. Apply the "summary" and "table" functions. Do a boxplot of all of the columns of the variables on the same plot.

4. Consider the following R code and resulting analysis:

```
> data(Titanic)  
> titanic <- as.data.frame(Titanic)  
> xtabs(Freq ~ Sex + Survived, data=titanic)
```

	Survived	
Sex	No	Yes
Male	1364	367
Female	126	344

```
> summary(xtabs(Freq ~ Sex + Survived, data=titanic))
```

```
Call: xtabs(formula = Freq ~ Sex + Survived, data = titanic)
```

```
Number of cases in table: 2201
```

Number of factors: 2

Test for independence of all factors:

$$\text{Chisq} = 456.9, \text{ df} = 1, \text{ p-value} = 2.302\text{e-}101$$

Use the evidence here to decide whether there is a male/female effect in survival (perhaps a chivalry test?). State your hypothesis, give the exact steps for testing this hypothesis, make a decision, and explain the evidence for your decision.

5. The European Union (EU) produced 1998 data for the (then) 15 member countries with two variables:
- (a) the median (EU standardized) income of individuals age 65 and older as a percentage of the population age 0–64,
 - (b) the percentage with income below 60% of the median (EU standardized) income of the national population.

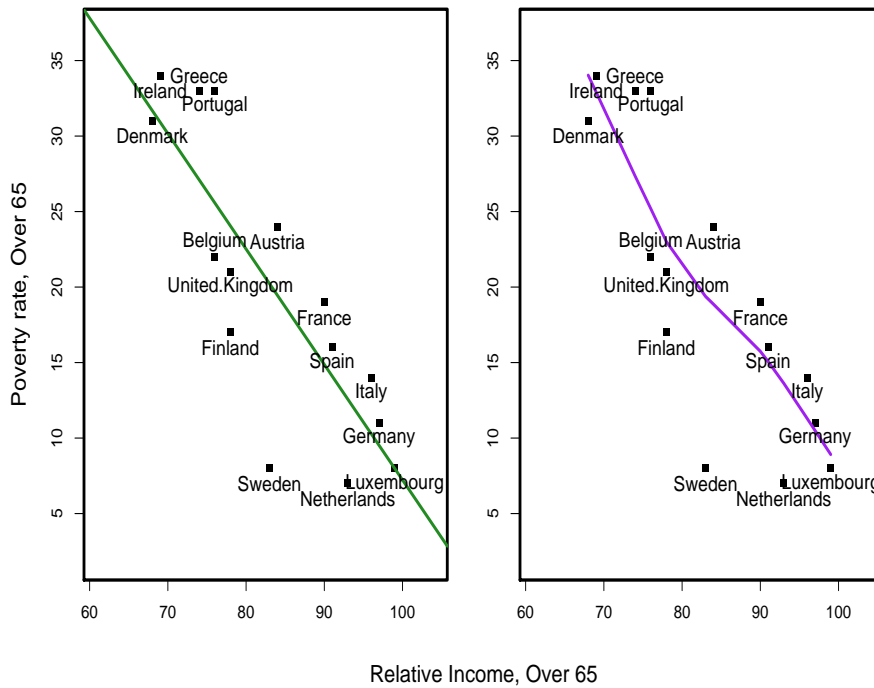
The data from the European Household Community Panel Survey are:

Nation	Relative Income	Poverty Rate
Netherlands	93.00	7.00
Luxembourg	99.00	8.00
Sweden	83.00	8.00
Germany	97.00	11.00
Italy	96.00	14.00
Spain	91.00	16.00
Finland	78.00	17.00
France	90.00	19.00
United.Kingdom	78.00	21.00
Belgium	76.00	22.00
Austria	84.00	24.00
Denmark	68.00	31.00
Portugal	76.00	33.00
Greece	74.00	33.00
Ireland	69.00	34.00

You can download and condition the data with the following three statements:

```
eu.pov <- read.table("http://people.hmdc.harvard.edu/~jgill/inc.pov.dat", row.names=1)
names(eu.pov) <- c("relative income", "poverty rate")
eu.pov <- eu.pov[-1,]
```

Run a linear model and Lowess smoother on these data and summarize the linear model fit. Produce a graph like the one below demonstrating the two models. Does the Lowess smoother help you defend or refute the linear fit?



6. Consider the bivariate normal PDF:

$$f(x_1, x_2) = \left(2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}\right)^{-1} \times \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)\right].$$

for $-\infty < \mu_1, \mu_2 < \infty$, $\sigma_1, \sigma_2 > 0$, and $\rho \in [-1 : 1]$.

For $\mu_1 = 3, \mu_2 = 2, \sigma_1 = 0.5, \sigma_2 = 1.5, \rho = 0.75$, calculate a grid search using R for the mode of this bivariate distribution on \mathbb{R}^2 . A grid search bins the parameter space into equal space intervals on each axis and then systematically evaluates each resulting subspace. First set up a two dimensional coordinate system stored in a matrix covering 99% of the support of this bivariate density, then do a systematic analysis of the density to show the mode without using “for” loops. Hint: see the R help menu for the outer function. Use the contour function to make a figure depicting bivariate contour lines at 0.05, 0.1, 0.2, and 0.3 levels.

7. Describe how to produce the chi-square and exponential distribution as special cases of the gamma distribution:

$$\mathcal{G}(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp[-x\beta], \quad 0 \leq x < \infty, \quad 0 < \alpha, \beta$$

8. Explain specifically how the Dirichlet distribution,

$$\mathcal{D}(\mathbf{x}|\alpha_1, \dots, \alpha_k) = \frac{\Gamma(\alpha_1 + \dots + \alpha_k)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_k)} x_1^{\alpha_1-1} \dots x_k^{\alpha_k-1} \quad 0 \leq x_i \leq 1, \sum_{i=1}^k x_i = 1, 0 < \alpha_i, \forall i \in [1, 2, \dots, k],$$

generalizes the beta distribution.