1 Lecture 5 Friday 01/26/01

Homework and Sectionning: see the logictics section

Here are some confidence interval applets to try:
Phil Stark's Webster West's SRS

7.3 Review of last time, if you missed see: Lecture 4

$$P(z_{\alpha/2} \leq \frac{\bar{X}-\mu}{s_{\bar{X}}} \leq z_{1-\alpha/2}) \doteq 1-\alpha$$

8 Estimating a Ratio

Suppose now that we have a population of N pairs $(x_1, y_1), (x_2, y_2) \ldots (x_N, y_N)$ The ratio of interest is

$$r = \frac{\sum_{i=1}^N y_i}{\sum_{i=1}^N x_i} = \frac{\mu_y}{\mu_x}$$

Motivating Examples :

1. Proportion of unemployed males of age 20-30 to number of males of age 20-30 in each household.

2. y = weekly expenditure in food for the household, x = number of inhabitants per household. r = weekly cost per inhabitant.

3. $y = \#$ motor vehicles in household
   $x = \#$ inhabitants of driving age
   r = mv per inhabitant of driving age

4. $y = \#$ acres of wheat
   $x = \#$ acres of plantation

We are going to do two things : Estimate a ratio,
Use this as a technique for estimating a mean μ_y.
First important remark:

$$r \neq \frac{1}{N}\sum_{i=1}^N \frac{y_i}{x_i}$$

The sample consists in random points (X_i, Y_i) the natural estimate for r is

$$R = \frac{\bar{Y}}{\bar{X}}$$

So R is a random variable, we will compute its expectation and variance, but R is NOT a linear function of $\bar{X}$ and $\bar{Y}$, we will need to use an approximation method commonly called the delta method or the method of error-propagation.

This is actually in section 4.6 of the book, so it may only be a revision for some of you, however for the benefit of others I am going to detail it, also because I have a colleague who says the only Maths in statistics is Taylor's theorem, he's wrong but it is fundamental.

8.1 Approximation Methods

In a situation when we only know the mean and the variance, not the whole sampling distribution of some random variable X, we are interested in the mean and the variance of some function of X, call it $Y = g(X)$.

We may want var(Y) to assess accuracy of an estiamte or just it's expectation.

If g is linear as we know finding it's expectation is trivial, and we can also find it's variance (recalling that $E(a + bV) = a + bE(V)$ and $Var(a + bV) = b^2Var(V)$)

So when g is not linear we have to bring ourselves back to that case by using a Taylor expansion, we hope to find a region where X dwells with high probability and where g is linear.

$$Y = g(X) \sim g(\mu_X) + (X - \mu_X)g'(\mu_X) + \frac{1}{2}(X - \mu_X)^2 g''(\mu_X)$$

This gives as a first order approximation to the expectation of Y :

$$E(Y) = g(\mu_X)$$

and as a second order approximation

$$E(Y) = g(\mu_X) + \frac{1}{2}\sigma_X^2 g''(\mu_X)$$

How good is such an approximation?

Depends on two things:

1. How nonlinear g is in a neighborhood of μ_X
2. How big σ_X^2 is.

We are just interested in the neighborhood of μ_X because that's where X lies most of the time,

Reminder : Chebychev's Theorem : $\mathrm{P}(|X - \mu_X| > k\sigma_X) < \frac{1}{k^2}$

Now we actually need a more sophisticated version, supposing we are interested in

$$Z = g(X, Y)$$

a function of two random variables we know about, we need two-dimensional Taylor series, note $\mu = (\mu_X, \mu_Y)$ Then

$$Z = g(X,Y) = g(\mu) + (X - \mu_X)\frac{\partial g}{\partial x}(\mu) + (Y - \mu_Y)\frac{\partial g}{\partial y}(\mu)$$

as a first approximation, this gives us straight away for the expectation $E(Z) = g(\mu)$ and with a little more work for the variance we have :

$$Var(Z) = \frac{\partial g}{\partial x}(\mu)^2\sigma_X^2 + \frac{\partial g}{\partial y}(\mu)^2\sigma_Y^2 + 2\frac{\partial g}{\partial x}(\mu)\frac{\partial g}{\partial y}(\mu)\sigma_{X,Y}$$

An improved estimate of the expectation can be obtained by taking the second order Taylor expansion:

$$\begin{aligned} Z = g(X,Y) = \quad & g(\mu) + (X - \mu_X)\tfrac{\partial g}{\partial x}(\mu) + (Y - \mu_Y)\tfrac{\partial g}{\partial y}(\mu) \\ & + \tfrac{1}{2}(X - \mu_X)^2\tfrac{\partial^2 g}{\partial x^2}(\mu) + \tfrac{1}{2}(Y - \mu_Y)^2\tfrac{\partial^2 g}{\partial y^2}(\mu) \\ & + (X - \mu_X)(Y - \mu_Y)\tfrac{\partial^2 g}{\partial x \partial y}(\mu) \end{aligned}$$

From this and from the properties of the expectation:

$$E(Z) = g(\mu) + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(\mu)\sigma_X^2 + \frac{1}{2}\frac{\partial^2 g}{\partial y^2}(\mu)\sigma_Y^2 + \sigma_{XY}\frac{\partial^2 g}{\partial x \partial y}(\mu)$$

For n-variables it is similar.

Note :

If you don't remember about Taylor series you can get the computer to remind you by using `Maple`.

'12