Properties of Estimators

Quantitative Methods II for Political Science
Kenneth Benoit

January 7, 2009

Objectives and learning outcomes

- to more deeply understand the linear regression model
- to diagnose and correct problems with LRM in real data
- to apply generalizations of LRM to binary and count data
- to be able to read quantitative studies in political science
- to know where to go for more advanced techniques and problems
- only prerequisite is Quant I, but the more previous statistics, the better

Grading

- Problem sets (50%)
  - Handed out Wed, due back the next Wed.
  - Problem solutions should be submitted as a single pdf file
  - Problem sets must be submitted to `http://turnitin.com`
  - You can scan anything using the departmental scanner if needed
- Replication project (50%) – quantative reanalysis of a published quantitative work

Texts and Software for this course

- Two primary texts
  - Kennedy, Peter. 1998. *A Guide to Econometrics*. 5th ed. Oxford: Blackwell
  - Julian J. Faraway. 2005. *Linear Models with R*. Boca Raton: Chapman & Hall.
  - The website for the Faraway book is `http://www.stat.lsa.umich.edu/~faraway/LMR/`
  - Can be ordered through Amazon or purchased at Hodges and Figgis (purchasing both is definitely recommended)

- Software will be the R statistical package
  - Free from `http://www.r-project.org`
  - Multi-platform
  - Extremely powerful
  - Lots of free documentation
  - There is a GUI shell called R Commander by John Fox, from `http://socserv.mcmaster.ca/jfox/Misc/Rcmdr/`

A problem: Method of Moments Failure

- Dublin uses serial numbers for cars such that 07-D-12371 means the year 2007, **D**ublin, and the 12371*th* registration number issued
- Let's say that based on a sample of observing number plates, you want to estimate the total number of licenses issued in 2007
- Sample: `12371, 5740, 432, 21999, 7629, 9000`
- The question is same as asking: What is N ?
- This is a version of a very common problem of estimating an equation for averages or the mean

- From the sample, we can calculate a sample mean $\bar{X}$: 9,528.5
- We also know that from the population of serial numbers $1, 2, 3, \ldots N$, the mean μ in terms of N is $\mu = (N+1)/2$
- If $E(\mu) = \bar{X}$, we can use this to solve for N:

$$\begin{aligned} \mu &= (N+1)/2 \qquad (1) \\ 2\mu &= N+1 \\ 2\mu - 1 &= N \\ N &= 2\bar{X} - 1 \\ &= 2(9,528.5) - 1 \\ &= 19,056 \end{aligned}$$

- So is answer 19,056?
- NO, since in this case we know it should be (at least) 21,999
- Lesson: Some methods of estimation are better than others!

Some suggestions at this point

- Suggestion: Review the Greek math alphabet, see `http://math.boisestate.edu/~tconklin/MATH144/Main/Extras/PRGreekAlphabet.pdf`
- Suggestion: Review the rules of matrix algebra
- Suggestion: Review the rules concerning expectations (and variances)

The Birthday Problem

- The “birthday problem” is a classic problem: Given a group of people, what is the probability of two people in the group having the same birthday?
- Approach problems like this first by considering extreme cases:
  - With one person, Pr()=0
  - With two persons, probability is very small
  - For a group of $\geq$ 365 people, Pr()=1 that two people will have the same birthday, since there are only 365 possible birthdays to go around
- Question: How many people are needed before the probability exceeds 0.50?

- For one person, there are 365 distinct birthdays (excluding leap years for simplicity)
- For two people:
  - there are 365^2 different pairs of birthdays for our two people
  - there are (365-1) different ways for the second person to have a different birthday
  - so the $\Pr(x_1 \neq x_2) = \frac{365(365-1)}{365^2}$
- A third person has 363 different days that could be different from the first two people, so $\Pr(x_1 \neq x_2 \neq x_3) = (365 \cdot 364 \cdot 363)/365^3$
- This leads to the general formula for the probability of a match with n birthdays being

$$\begin{aligned} Pr(n) &= 1 - \frac{365^n - 365 \cdot 364 \cdot \ldots \cdot (365 - n + 1)}{365^n} \\ &= 1 - \frac{365!}{(365-n)!365^n} \end{aligned} \quad (2)$$

Randomness and statistical modelling

- The disturbance term: $Y = f(X) + \epsilon$. The ϵ makes the function *stochastic*; without it the function would be *deterministic*.
- Where does ϵ come from?
  1. Omission of the influence of innumerable chance events.
  2. Measurement error.
  3. Human indeterminacy.
- Parameter is generally β or θ. Estimates will be $\hat{\beta}$ or $\hat{\theta}$.
- Most common estimation method is to minimize the squared errors ("least squares"). What are alternatives? (1) absolute deviations, (2) horizontal deviations, (3) etc.

Point Estimation

- How can the population be estimated from the sample?
- A random sample is a random subset of the population
- "Strictly random" means all units from the population have an equal probability of being chosen for the sample being chosen for the sample

Sample	**Population**
Relative frequencies $\frac{f_i}{n}$ are used to compute: Sample mean $\bar{X}$ Sample variance s^2	Probabilities are used to compute Population mean μ Population variance σ^2
These random variables are *statistics* or estimators	These fixed constants are *parameters* or targets

Table: Review of Population v. Sample

Properties of Estimators: Bias

- U is an unbiased estimator of θ if $E(U) = \theta$. An estimator V is called biased if $E(V)$ is different from θ
- Bias $\equiv E(V) - \theta$
- Bias is often assessed by characterizing the sampling distribution of an estimator
    - *repeated* samples are drawn by resampling from the disturbance term (in our case, ϵ), while keeping the values of the independent variables unchanged
    - For instance we could do this 1,000 times using β^* to calculate an estimate of β
    - The way that the 1,000 samples are distributed is called the sampling distribution of β^*
    - For an estimator β^* to be an unbiased estimator of β means that the mean of its sampling distribution is equal to β
    - Another way to put this is that $\mathrm{E}(\beta^*) = \beta$

Properties of Estimators: Efficiency

- We would like the distribution of an estimator to be highly concentrated—to have a small variance. This is the notion of *efficiency*. The efficiency of V compared to W is $W \equiv \frac{\mathrm{var}\, W}{\mathrm{var}\, V}$.
- If population being sampled is exactly symmetric, then center can be estimated without bias by either the sample mean $\bar{X}$ or the sample median X'. For large samples, $\mathrm{var} X' \approx 1.57\sigma^2/n$. Since $\bar{X}$ has variance σ^2/n, the smaller variance makes it 157% more efficient than the median for normal populations.
- This gives rise to the notion of *relative efficiency*, to which we will return shortly
- Not really the same as "minimum variance"

Properties of Estimators: Consistency

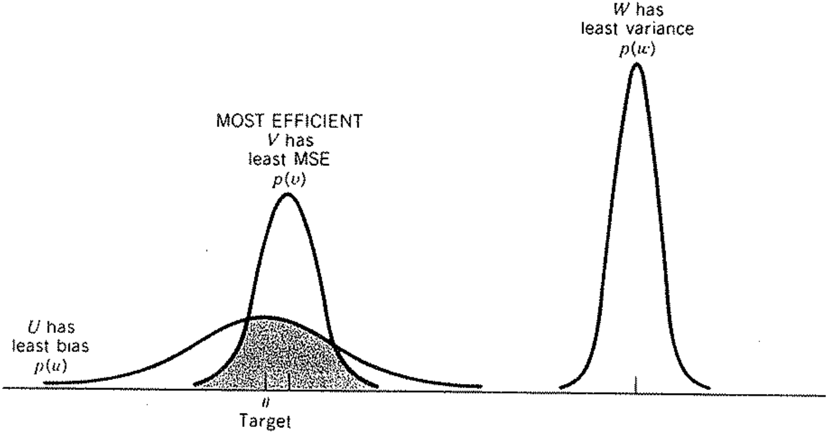
- A consistent estimator is one that concentrates in a narrower and narrower band around its target as sample size increases indefinitely. MSE approaches zero in the limit: bias *and* variance both approach zero as sample size increases.
- V is defined to be a consistent estimator of θ, if for any positive δ (no matter how small), $\Pr(|V - \theta|) < \delta) \longrightarrow 1$, as $n \longrightarrow \infty$
- (Kennedy) If the asymptotic distribution of $\hat{\beta}$ becomes concentrated on a particular value k as $N \longrightarrow \infty$, k is said to be the *probability limit* of $\hat{\beta}$ and is written $\text{plim}\hat{\beta} = k$; if $\text{plim}\hat{\beta} = \beta$, then $\hat{\beta}$ is said to be *consistent*

Choosing from among alternative estimators

- When we compare two *unbiased* estimators, which should we choose?
- Answer: The one with **minimum variance**
- When comparing both biased, and unbiased, which should we choose?
- Answer: The one with the best combination of small bias and small variance
- **Mean Squared Error (MSE):** $\equiv E(V-\theta)^2$.

More on mean squared error

- MSE $=$ (variance of estimator) $+$ (its bias)^2



- Relative efficiency of V compared to W: $\equiv \frac{\mathrm{MSE}(W)}{\mathrm{MSE}(V)}$

Large-sample properties of estimators

- *asymptotically unbiased*: means that a biased estimator has a bias that tends to zero as sample size approaches infinity.
- When no estimator with desireable small-scale properties can be found, we often must choose between different estimators on the basis of *asymptotic* properties
- Asymptotic properties of estimators refer to what happens as sample size increases towards infinity
- Many estimators are trusted in principle because of their asymptotic properties, even when these don't hold in smaller samples (e.g. maximum likelihood)
- For many estimation problems, non-parametric alternatives are favored when sample sizes are small
  - Example: t-test versus Kruskal Wallis test; or Chi-squared test versus Fisher exact test

Example: mean squared deviation

- Mean squared deviation or $MSD = \frac{1}{n} \sum (X - \bar{X})^2$
- This is a biased estimator of population variance σ^2, since on average it will underestimate true quantity
- For example, when $X = 1$, it yields MSD $= 0$
- As a result, we use instead the *sample variance*:

$$s^2 \equiv \frac{1}{n-1} \sum (X - \bar{X})^2$$

- But MSD is *asymptotically unbiased*, since its bias approaches zero as $n \rightarrow \infty$

Proof

$$\begin{aligned} \mathrm{MSD} &= \left(\frac{n-1}{n}\right) s^2 && (3)\\ &= \left(1-\frac{1}{n}\right) s^2 && (4)\\ E(\mathrm{MSD}) &= \left(1-\frac{1}{n}\right) E(s^2) && (5) \end{aligned}$$

Since s^2 is an unbiased estimator of σ^2:

$$\begin{aligned} E(\mathrm{MSD}) &= \left(1-\frac{1}{n}\right) E(\sigma^2) && (6)\\ &= \sigma^2 - \left(\frac{1}{n}\right) \sigma^2 && (7) \end{aligned}$$

The last term $\left(\frac{1}{n}\right) \to 0$ as $n \to \infty$.

Maximum likelihood (very brief introduction)

- Based on the principle that the sample of data at hand is more likely to have come from a world characterized by one particular set of parameter values, than from any other set of values
- Example: Given a set of coin toss data, what is the value of π (the probability that $x_i =$head) that is most likely to have generated the data?
- Properties:
  - asymptotically unbiased
  - consistent
  - (asymptotically) normally distributed
  - asymptotic variance can be computed using a standard formula
- (almost all) maximization of likelihoods is done numerically using computers
- The logit, probit, Poisson etc. models we will do later in this class all use maximum likelihood for estimating parameters

Monte Carlo studies

- A *Monte Carlo* study is a simulation exercise designed to shed light on the small-sample properties of competing estimators for a given estimation problem
- Used when small-sample properties cannot be derived theoretically, or as a supplement to theoretical derivations
- Allows direct exploration of samplong distributions, through simulation
- Steps involved:
  1. Model the data-generating process
  2. Genereate artificial datasets
  3. create estimates from the data using the estimator
  4. use these estimates to assess the estimator's sampling distribution
- Monte Carlo simulation is extremely common and important tool of modern statistical methods, and computationally very accessible using modern computers and software (like R)

Monte Carlo example

- Consider the sample variance estimator $s^2 = \frac{n}{n-1}\bar{y}^2$
- Cochran's theorem shows that if Y is *iid* Normal, then s^2 follows a scaled chi-square distribution χ^2_{n-1}
- To verify this using Monte Carlo simulations, we can construct sample datasets and examine the sampling distribution, for a given sample size

```
# define a function for the sample variance
sv <- function(y) { length(y) / (length(y)-1) * mean(y)^2 }
# create a loop to compute and store 1,000 simuated sample variances
# this will be from 50 random y values
result <- numeric(500)
for (i in 1:500) { result[i] <- sv(rnorm(50)) }
# plot the result, after sorting the computed sample variances
plot(sort(result), type="l", ylab="Computed sample variance")
# now check whether sampling distribution matches a Chi^2
chisq.test(result)
```

Working in R: Birthday problem example

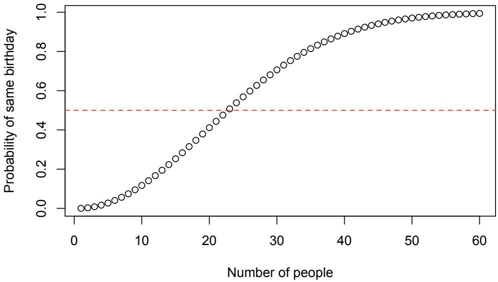
- Formula: $1 - \frac{365!}{(365-n)!365^n}$
- In R, w can use the `factorial()` function
- So for $n = 10$:
  `1 - (factorial(365) / (factorial(365-n) * 365`^{n}`))`
- Does this work? No – numbers too big!
- How to solve this: use logarithms and `lfactorial()`:
  `1-exp(lfactorial(365) - lfactorial(365-n) - n*log(365))`

Working in R: Birthday problem example code

```
lbdp <- function(n) {
  1 - exp(lfactorial(365) - lfactorial(365-n) - n*log(365))
}

x <- 1:60
plot(x,lbdp(x))

plot(x,lbdp(x),
     xlab="Number of people",ylab="Probability of same birthday")
abline(h=.5, lty="dashed", col="red")
```



Example: Regression output

Valid cases:	4274	Dependent variable:	disprls
Missing cases:	0	Deletion method:	None
Total SS:	2094312.971	Degrees of freedom:	4251
R-squared:	0.887	Rbar-squared:	0.886
Residual SS:	237366.238	Std error of est:	7.472
F(22,4251):	1511.642	Probability of F:	0.000

Variable	Estimate	Standard Error	t-value	Prob >|t|	Standardized Estimate	Cor with Dep Var
CONSTANT	50.437041	0.164518	306.573980	0.000	---	---
HSL*m	-8.501692	2.525842	-3.365885	0.001	-0.082936	-0.215230
HSL	-34.443131	2.579062	-13.354906	0.000	-0.329397	-0.216392
SL*m	-6.526475	2.525842	-2.583881	0.010	-0.063667	-0.223110
SL	-37.302552	2.579062	-14.463611	0.000	-0.356743	-0.225317
MSL*m	-7.828347	2.525842	-3.099302	0.002	-0.076367	-0.217458
MSL	-35.371193	2.579062	-13.714750	0.000	-0.338273	-0.218966
dH*m	-8.292628	2.525842	-3.283115	0.001	-0.080896	-0.207012
dH	-33.823319	2.579062	-13.114581	0.000	-0.323470	-0.208080
LRH*m	-6.953863	2.525842	-2.753087	0.006	-0.067836	-0.224528
LRH	-37.002049	2.579062	-14.347095	0.000	-0.353869	-0.226579
LRDr*m	-7.023068	2.525842	-2.780486	0.005	-0.068511	-0.222815
LRDr	-36.755473	2.579062	-14.251488	0.000	-0.351511	-0.224798
LRI*m	-7.679571	2.525842	-3.040401	0.002	-0.074916	-0.217981
LRI	-35.579349	2.579062	-13.795460	0.000	-0.340263	-0.219566
ImpHA*m	-10.721835	2.525842	-4.244856	0.000	-0.104594	-0.157791
ImpHA	-26.278325	2.579062	-10.189101	0.000	-0.251313	-0.156677
EqP*m	-21.029154	2.525842	-8.325603	0.000	-0.205144	-0.147518
EqP	-14.500895	2.579062	-5.622546	0.000	-0.138679	-0.141654
Dan*m	-7.355209	2.525842	-2.911983	0.004	-0.071752	-0.220301
Dan	-36.153527	2.579062	-14.018091	0.000	-0.345755	-0.222081
Ad*m	-20.961184	2.525842	-8.298693	0.000	-0.204481	-0.145850
Ad	-14.401702	2.579062	-5.584085	0.000	-0.137731	-0.139978