Appendix C
Fundamentals of Mathematical Statistics

Jim Bang

Approaches to Parameter Estimation

Ordinary Least Squares (OLS)

$$\min_{\hat{\beta}} \sum_{i=1}^{n} (Y_i \,\breve{}\, \beta_0 \,\breve{}\, \beta_1 X_i)^2$$

Least/Minimum Absolute Deviations (LAD/MAD)

$$\min_{\hat{\beta}} \sum_{i=1}^{n} |Y_i \,\breve{}\, \beta_0 \,\breve{}\, \beta_1 X_i|$$

Maximum likelihood (MLE)

$$\max_{\hat{\beta}} \prod_{i=1}^{n} f(Y_i \,\breve{}\, \beta_0 \,\breve{}\, \beta_1 X_i)$$

where $f(\cdot)$ is the probability distribution of the errors (e.g. Normal, logistic, Poisson)

Method of Moments

$$E(Xu) = 0$$

$$\sum_{i=1}^{n} X_i (Y_i - \beta_0 - \beta_1 X_i) = 0$$

If the assumptions of the Classical Regression Model hold, all four of these methods are equivalent.

Interval Estimation

The 95% confidence interval for the sample mean, $\bar{x}$ solves

$$P(\bar{x} - t^c_{0.025} \cdot s_{\bar{x}} \leq \mu \leq \bar{x} + t^c_{0.025} \cdot s_{\bar{x}}) = 0.95$$

Since the sampling distribution of $\bar{x}$ is normal and the sampling distribution of $s^2_{\bar{x}}$ is χ^2, this involves inverting the t-distribution (normal divided by the square root of χ^2).

Example

Using the *audit* data from the *wooldridge*, calculate the 95% confidence interval *by hand* by calculating the following (and naming them):

1. the mean of `y` (call it `avgy`);
2. the number of observations in (`length()` of) `y` (`n`);
3. the standard deviation of `y` (`sdy`)
4. the standard error of $\bar{y}$ (`sdy`)
5. the two-sided critical value for $\alpha = 0.05$ (`c05`)
6. the 95% confidence-interval bounds as a vector (lower bound, upper bound)

```
avgy<- mean(audit$y)
n   <- length(audit$y)
sdy <- sd(audit$y)
se  <- sdy/sqrt(n)
c05   <- abs(qt(.025, n-1))
avgy + c05 * c(-se,+se)
```

```
## [1] -0.1939385 -0.0716217
```

Hypothesis Testing

1. State the Null & Alternative Hypotheses
2. Determine the Significance Level (α)
3. Calculate the parameters and the test statistic
4. Calculate the critical value or p-value
5. Make a Rejection Decision

Example

Calculate the test statistic, absolute critical value for $\alpha = 0.01$, and p-value to test the null hypothesis of $H_0 : \mu_y = 0$ against the two-sided alternative of $H_1 : \mu_y \neq 0$. Call these t, c01, and p, and print them as a vector.

```
t <- avgy/se
c01 <- abs(qt(0.005, n-1))
p <- pt(t, n-1)
c(t, c01, p)
```

```
## [1] -4.276816e+00  2.596469e+00  1.369271e-05
```

Shortcut using *t.test*()

Replicate the confidence interval and t-test from the previous examples using the `t.test()` command. Don't forget that the second example used a different α than the first one.

```
t.test(audit$y)
```

```
##
##  One Sample t-test
##
## data:  audit$y
## t = -4.2768, df = 240, p-value = 2.739e-05
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.1939385 -0.0716217
## sample estimates:
##  mean of x
## -0.1327801
```

```
t.test(audit$y, conf.level = 0.99)
```

```
##
```

```
## 	One Sample t-test
## 
## data:  audit$y
## t = -4.2768, df = 240, p-value = 2.739e-05
## alternative hypothesis: true mean is not equal to 0
## 99 percent confidence interval:
##  -0.21339131 -0.05216886
## sample estimates:
##  mean of x 
## -0.1327801
```