

Supplementary Notes for Homework 2, Exercise 2

Advanced Quantitative Methods (PLSC 504)

Fall 2017

Exercise 2 Notes

This is a supplement to the Exercise 2 solutions guide posted on Canvass. We computed analytically that both \hat{Y}_s and \hat{Y}_d are unbiased. We can confirm this with simulation:

```
# Make up some values for Y1-Y3
Y1 <- 3
Y2 <- 4
Y3 <- 5

# Calculate our target parameter
Y_Avg <- (Y1 + Y2 + Y3)/3

N <- 10^4
Y_s <- NA
Y_d <- NA

# Run a simulation.
set.seed(504)
for(i in 1:N){
  tmp <- sample(c(Y1, Y2, Y3), 2, replace = FALSE)
  Y_s[i] <- tmp[1]
  Y_d[i] <- mean(tmp)
}

# Bias:
mean(Y_s) - Y_Avg
```

```
## [1] -0.003
```

```
mean(Y_d) - Y_Avg
```

```
## [1] 0.00165
```

The simulation confirms what we derived, so we can feel more confident that we were correct in asserting that both estimators are unbiased. Next, we check the variances for each estimator. You didn't need to simplify the math for the variance of either estimator. Indeed it was quite tedious. But incase you did, we can confirm that what we claim makes sense using the simulated results:

```
# Variance of Y_s
var(Y_s)
```

```
## [1] 0.6634573
```

```
# Simple expression
```

```
(Y1^2 + Y2^2 + Y3^2)/3 - (Y_Avg)^2
```

```
## [1] 0.6666667
```

```
# After tedious algebra
```

```
(2*Y1^2+2*Y2^2+2*Y3^2-2*Y1*Y2-2*Y1*Y3-2*Y2*Y3)/9
```

```
## [1] 0.6666667
```

```
# Variance of  $Y_d$   
var(Y_d)
```

```
## [1] 0.1665889
```

```
# Simple expression  
((Y1 + Y2)^2 + (Y1 + Y3)^2 + (Y2 + Y3)^2)/12 - (Y_Avg)^2
```

```
## [1] 0.1666667
```

```
# After tedious algebra  
(2*Y1^2 + 2*Y2^2 + 2*Y3^2 - 2*Y1*Y2 - 2*Y1*Y3 - 2*Y2*Y3)/36
```

```
## [1] 0.1666667
```

As expected, $\text{Var}(\hat{Y}_s) = 4 \cdot \text{Var}(\hat{Y}_d)$. We can also work out the p -value calculations in R step by step. The first step for either estimator is to *assume* that the null hypothesis we are given is true:

```
# P-values.  
# Step 1: assume Null is true.
```

```
Y1 <- 1
```

```
Y2 <- 2
```

```
Y3 <- 2
```

```
# Calculate target parameter under the null
```

```
Y_Avg <- (Y1 + Y2 + Y3)/3
```

```
Y_Avg
```

```
## [1] 1.666667
```

Now, start with \hat{Y}_s and calculate the distribution of the estimator and the error distribution (e.g. $\hat{Y}_s - Y_{AVG}$) under the null.

```
# Distribution of  $Y_s$  under Null. We are twice as likely to draw  $Y_a = 2$ .
```

```
Y_s <- c(1, 2, 2)
```

```
# Distribution of error under Null.
```

```
e_s <- Y_s - Y_Avg
```

```
e_s
```

```
## [1] -0.6666667 0.3333333 0.3333333
```

So we conclude that $\Pr(\epsilon_s = -2/3) = 1/3$ and $\Pr(\epsilon_s = 1/3) = 2/3$. Now we are told that we observed an estimate of $\hat{Y}_s = 1$. So we calculate the error associated this estimate (again assuming the Null!) and the associated p -value:

```
# Error when  $Y_s = 1$ ?
```

```
1 - Y_Avg
```

```
## [1] -0.6666667
```

```
# P.Value: What proportion of  $e_s$  have same or larger magnitude as the  
# observed error?
```

```
mean(abs(e_s) >= abs(1 - Y_Avg))
```

```
## [1] 0.3333333
```

Now we repeat for \hat{Y}_d :

```

# Now for Y_d
Y_d <- c(3/2, 3/2, 4/2)

# Distribution of error under Null.
e_d <- Y_d - Y_Avg

# Error when Y_d = 1?
1 - Y_Avg

## [1] -0.6666667

# P.Value: What proportion of e_d have same or larger magnitude as the
# observed error?
mean(abs(e_d) >= abs(1 - Y_Avg))

## [1] 0

```