ECN 102: Analysis of Economics Data

Homework 2 Answer Key

Remy Beauregard

Question 1: Population mean and variance

Let X denote the number of days it will rain in Davis, CA in July. Suppose $X = 0$ with probability 0.75, $X = 1$ with probability 0.13, $X = 2$ with probability 0.08, and $X = 3$ with probability 0.04 (assume these are all the possible values X can take).

a) Obtain the population mean, $\mu = E[X]$, from first principles

$\mu = 0 \times 0.75 + 1 \times 0.13 + 2 \times 0.08 + 3 \times 0.04 = 0.41$

b) Obtain $\sigma^2 = E[(x - \mu)^2]$ from first principles

$\sigma^2 = (0 - 0.41)^2 \times 0.75 + (1 - 0.41)^2 \times 0.13 + (2 - 0.41)^2 \times 0.08 + (3 - 0.41)^2 \times 0.04) = 0.64$

c) Find the standard deviation of X

$\sigma = \sqrt{\sigma^2} = \sqrt{0.64} = 0.8$

d) What do these Greek letters μ and σ correspond to, specifically? How do these differ from $\bar{x}$ and s? **Mu is our population mean and sigma is our population standard deviation. X-bar and s are our sample statistic counterparts.**

Question 2: Mean and variance of the sample mean

Let $\bar{X}$ be the mean of a random sample of size $n = 400$ from a random variable X that is *not* distributed normally, with mean 45 and variance 100.

a) Give the mean of $\bar{X}$ $E[\bar{X}] = \mu = 45$

b) Give the variance and standard deviation of $\bar{X}$ $V[\bar{X}] = \frac{\sigma^2}{n} = \frac{100}{400} = \frac{1}{4}, SD[\bar{X}] = \sqrt{\frac{1}{4}} = \frac{1}{2}$

c) Is X normally distributed? Is $\bar{X}$ likely to be normally distributed? Explain. **X is not normally distributed, given in the question. However, due to the CLT with $n = 400 > 30$, $\bar{X}$ is expected to be normally distributed, such that $\bar{X} \sim N(45, \frac{1}{4})$.**

Question 3: The t distribution

For a given T distribution, the following Stata commands give areas in the **right tail** of the distribution:

- `di ttail(9,1.2)` gives $P[T > 1.2]$ for $T \sim T(9)$ (a t-distribution with $n = 10, df = 9$).
- `di invttail(9,0.025)` gives $t^* : P[T > t^*] = 0.025$ again for $T \sim T(9)$

Now, for $T \sim T(44)$, use these Stata commands to find:

a) $P[T > 2]$ **`di ttail(44,2)`**

b) $P[|T| > 2]$ **`di ttail(44,2)*2`**

c) $P[T < -1.8]$ **`di ttail(44,1.8)` or `di 1-ttail(44,-1.8)`; distribution is symmetric**

d) $P[T < 1.8]$ **`di 1-ttail(44,1.8)` or `di ttail(44,-1.8)`; probabilities sum to 1**

e) $P[T > 0]$ **0.50; distribution is symmetric and centered at 0**

f) $t^* : P[T < -t^*] = 0.025$ **`di -invttail(44,0.025)` or `di invttail(44,1-0.025)`; symmetric**

g) $t^* : P[T > t^*] = 0.05$ **`di invttail(44,0.05)`**

h) $t^* : P[|T| > |t^*|] = P[T < -t^* \text{ or } T > t^*] = 0.05$ **`di invttail(44,0.05/2)`; 2-sided, symmetric**

i) $t^* : P[T > t^*] = 0.5$ **0; distribution is symmetric with 50% above or below zero**

Note: (e) and (i) should be answerable without access to Stata

Question 4: T-testing in Stata

Use AED_CALELECTRICITY.DTA from hw1. Suppose we want to test whether the spot price of electricity is different from its one-day ahead forward price.

a) Describe the variables present in the dataset. Which variable corresponds to the spot price? Which corresponds to the forward price? Is there a variable for the difference between the two?

```
use AED_CALELECTRICITY, clear
describe
```

```
(Data for A. Colin Cameron (2015): Analysis of Economics Data, W.W. Norton)

Contains data from AED_CALELECTRICITY.dta
 Observations:           682                  Data for A. Colin Cameron
                                                (2015): Analysis of Economics
                                                Data, W.W. Norton
    Variables:             7                  2 Mar 2015 20:17
--------------------------------------------------------------------------------
Variable      Storage   Display    Value
    name         type    format    label      Variable label
--------------------------------------------------------------------------------
month           byte    %8.0g                 Month of year number
year            int     %8.0g                 Year
day             byte    %8.0g                 Day of month
hour            byte    %8.0g                 Hour of day (24 hour clock)
npx             float   %9.0g                 One-day ahead forward price
                                                California ($/MWH)
niso            float   %9.0g                 Spot price California ($/MWH)
diff            float   %9.0g                 niso - npx
--------------------------------------------------------------------------------
Sorted by:
```

Spot price is *niso*, forward price is *npx*, difference is *diff*.

b) It is given above that we want to test whether the difference between these two series is different from zero. Is this a one-sided or two-sided test? What would our null and alternate hypotheses be?

$$\text{Two-sided test; } H_0 : \mu_{\text{diff}} = 0, H_A : \mu_{\text{diff}} \neq 0$$

c) To run a ttest on a variable in Stata, we use `ttest [VARNAME]=[NULL VALUE]`. Include the output Stata gives for running this command.

```
ttest diff=0
```

```
One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]
---------+--------------------------------------------------------------------
    diff |     681    .2296579    1.350936      35.254    -2.42285    2.882165
------------------------------------------------------------------------------
```

```
    mean = mean(diff)                                             t =   0.1700
H0: mean = 0                                     Degrees of freedom =      680

    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 0.5675         Pr(|T| > |t|) = 0.8651          Pr(T > t) = 0.4325
```

d) What is the conclusion of our test based on the p-value approach?

Our p-value for our 2-sided test is $0.8561 > \alpha = 0.05$**, meaning we** fail to reject** our null hypothesis. We have **insufficient** evidence to say that the spot price and forward price are different for this sample.**

e) What is the conclusion of our test based on the critical value approach? (You may need to compute an additional quantity to answer this)

```
di invttail(682-2,0.025)
```

```
1.9634587
```

Our t-stat of $0.17 < t^*_{0.025,680} = 1.96$ **meaning we also fail to reject our null with the critical value approach.**

f) What is the conclusion of our test based on the given confidence interval?

Our difference value of 0 assumed under the null *is* included in our 95% confidence interval, meaning we also fail to reject our null.

g) If we had instead wanted to test whether the difference between the spot price and forward price was *greater than zero*, what would our hypotheses have been? What would our conclusion have been?

$$\text{One-sided test; } H_0 : \mu_{\text{diff}} \leq 0, H_A : \mu_{\text{diff}} > 0$$

Our p-value is now $0.4325 > \alpha = 0.05$**, so we again fail to reject our null. We also have insufficient evidence to say that the difference between the forward and spot prices is greater than zero.**

h) Why does Stata provide three p-values and alternate hypotheses in its output for `ttest` but only one t-statistic?

Our computed t-statistic will be the same for all three tests, but our alternate hypothesis will change and thus the p-value associated with our statistic and/or critical value we compare our statistic to will also change.