Economics 102: Analysis of Economic Data
Cameron Spring 2015 April 23
Department of Economics, U.C.-Davis

First Midterm Exam (Version A)

Compulsory. Closed book. Total of 30 points and worth 22.5% of course grade.
Read question carefully so you answer the question.
You are to use only **simple calculations (+, -, /, *, square root) and show all workings**.
Do not use calculators with graphical or statistical analysis capabilities.
For computations final answers should be to at least four significant digits.
You may remove the formula sheet and the Stata output sheet(s) at end of exam.

Question scores

Question	1*a*	1*b*	1*c*	1*d*	1*e*	1*f*	2*a*	2*b*	2*c*	2*d*	3*a*	3*b*	3*c*	3*d*	*Mult.choice*
Points	1	1	1	2	2	2	3	3	1	1	2	3	2	1	5

QUESTIONS 1-2 USE STATA OUTPUT GIVEN AT THE END OF THIS EXAM. For some questions the answer is given directly in the output. For other questions you will need to use the output plus additional computation.

1. This question uses annual data on average tuition and fees at four-year public universities.
`growth` = Percentage annual growth in real tuition and fees

(a) Given the statistical output does `growth` appear to be normally distributed? **Explain.**

(b) What type of chart would you use to visually see whether `growth` is normally distributed?

(c) Give the interquartile range for variable `growth`.

(d) If real tuition and fees grow at the same average rate as it has from 1986 to 2014, in what future year will they be double their value in 2014? **Explain your answer.**

(e) How much has the overall price level, used to convert nominal values to real values, increased by from 1986 to 2014? **Explain your answer.**

(f) Given the summary statistics for variable `lnnominal` what is the approximate annual increase in nominal tuition and fees from 1986 to 2014? **Explain your answer.**

2. For this question continue to use the information given at the end of the exam. Clearly state any details that you use along the way.

(a) Give a **95 percent** confidence interval for the growth in real tuition and fees.

(b) The claim is made that the annual increase in real tuition and fees from 1986 to 2014 was five percent. Test this claim at significance **level 0.10**.

(c) Give the Stata code that will generate variable `growth` from variable `real.`

(d) The original data were in a file called `tuition.csv`. Give the Stata command you would use to read this data into Stata.

3.(a) Consider a simple random sample of size 4 with values $6, 4, 2, 8$.
Compute the sample mean and sample variance. **Show all workings.**

(b) Let X be the number of initial public offerings (IPO) in excess of $100 million in 2016. Suppose $X = 10$ with probability 0.1, $X = 20$ with probability 0.3 and $X = 30$ with probability 0.6.
Compute the mean, variance and standard deviation of X. **Show all workings.**

(c) Suppose for $X \sim (200, 10^2)$ we form 100 samples of size 50 and obtain 100 sample means $\bar{x}$.
What approximately do you expect the average of the $\bar{x}$ to equal?
What approximately do you expect the standard deviation of the $\bar{x}$ to equal?
Explain your answer.

(d) Given the following code what value do your expect for the sample mean of variable `x`?
Explain.

```
set obs 1000
set seed 10101
generate u = runiform()
hist u, start(0) width(0.1)
generate x = 0
replace x = 1 if u < 0.4
```

Multiple Choice Questions (1 point each)

1. Data on number of doctor visits in each year from 2000 to 2012 for a sample of 192 individuals is an example of

a. numerical cross-section data

b. numerical panel data

c. numerical time series data

d. none of the above

2. For a sample of size 93 on a categorical variable with five categories a useful visual summary is provided by

a. a pie chart

b. a box-and-whisker plot

c. a line chart

d. none of the above

3. For a population of 400 million there are 160 million employed and 40 million unemployed. The unemployment rate is

a. 25 percent

b. 20 percent

c. 15 percent

d. 10 percent

e. none of the above

4. A better picture of the overall U.S. stock market is provided by

a. Dow Jones Industrial Average

b. Standard and Poors 500 Index

c. Nasdaq Composite Index

5. The $T(n-1)$ distribution is used in statistical inference because

a. X is often $T(n-1)$ distributed

b. $\bar{X}$ is $T(n-1)$ distributed in large samples

c. both a. and b.

d. neither a. not b.

Cameron: Department of Economics, U.C.-Davis
SOME USEFUL FORMULAS

Univariate Data

$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ *and* $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

$\bar{x} \pm t_{\alpha/2;n-1} \times (s_x/\sqrt{n})$ *and* $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$

ttail$(df, t) = \Pr[T > t]$ where $T \sim t(df)$

$t_{\alpha/2}$ such that $\Pr[|T| > t_{\alpha/2}] = \alpha$ is calculated using invttail$(df, \alpha/2)$.

Bivariate Data

$r_{xy} = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \times \sum_{i=1}^{n}(y_i - \bar{y})^2}} = \dfrac{s_{xy}}{s_x \times s_y}$ [Here $s_{xx} = s_x^2$ and $s_{yy} = s_y^2$].

$\widehat{y} = b_1 + b_2 x_i$ $b_2 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$ $b_1 = \bar{y} - b_2\bar{x}$

TSS $= \sum_{i=1}^{n}(y_i - \bar{y}_i)^2$ ResidualSS $= \sum_{i=1}^{n}(y_i - \widehat{y}_i)^2$ Explained SS = TSS - Residual SS

$R^2 = 1-$ ResidualSS/TSS

$b_2 \pm t_{\alpha/2;n-2} \times s_{b_2}$

$t = \dfrac{b_2 - \beta_{20}}{s_{b_2}}$ $s_{b_2}^2 = \dfrac{s_e^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$ $s_e^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2$

$y|x = x^* \in b_1 + b_2 x^* \pm t_{\alpha/2;n-2} \times s_e \times \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_i (x_i - \bar{x})^2} + 1}$

$\mathrm{E}[y|x = x^*] \in b_1 + b_2 x^* \pm t_{\alpha/2;n-2} \times s_e \times \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}}$

Multivariate Data

$\widehat{y} = b_1 + b_2 x_{2i} + \cdots + b_k x_{ki}$

$R^2 = 1-$ ResidualSS/TSS $\bar{R}^2 = R^2 - \frac{k-1}{n-k}(1 - R^2)$

$b_j \pm t_{\alpha/2;n-k} \times s_{bj}$ *and* $t = \dfrac{b_j - \beta_{j0}}{s_{b_j}}$

$F = \dfrac{R^2/(k-1)}{(1 - R^2)/(n-k)} \sim F(k-1, n-k)$

$F = \dfrac{(\text{ResSS}_r - \text{ResSS}_u)/(k-g)}{\text{ResSS}_u/(n-k)} \sim F(k-g, n-k)$

Ftail$(df1, df2, f) = \Pr[F > f]$ where F is $F(df1, df2)$ distributed.

F_α such that $\Pr[F > f_\alpha] = \alpha$ is calculated using invFtail$(df1, df2, \alpha)$.

```
. describe

Contains data from mts15.dta
  obs:            29
 vars:             6                          20 Apr 2015 15:35
 size:           522
-------------------------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------------------------
year            int     %8.0g                 Year
nominal         int     %8.0g                 Tuition and fees at 4 year public (current $)
real            int     %8.0g                 Tuition and fees at 4 year public (constant 2014 $)
lnnominal       float   %9.0g                 Natural logarithm of variable nominal
lnreal          float   %9.0g                 Natural logarithm of variable real
growth          float   %9.0g                 Percentage annual growth in variable real
-------------------------------------------------------------------------------------------------
Sorted by: year

. summarize, sep(6)

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        year |         29        2000    8.514693       1986       2014
     nominal |         29    4184.172    2271.775       1259       8471
        real |         29    5321.379    1805.169       2739       8471
   lnnominal |         29     8.18714    .5753116   7.138073   9.044404
      lnreal |         29    8.523088    .3441964   7.915348   9.044404
      growth |         28    4.155236    2.974328  -.5972697   10.44776

. summarize growth, detail

          Percentage annual growth in variable real
-------------------------------------------------------------
      Percentiles      Smallest
 1%    -.5972697      -.5972697
 5%     .4553079       .4553079
10%     1.010702       1.010702       Obs                  28
25%     2.056625       1.037691       Sum of Wgt.          28

50%     3.288075                      Mean           4.155236
                        Largest       Std. Dev.      2.974328
75%     5.993356       8.543264
90%     8.990582       8.990582       Variance       8.846628
95%     10.00192       10.00192       Skewness       .6414501
99%     10.44776       10.44776       Kurtosis       2.458796

. display "t_27,.005 = " invttail(27,.005) "  t_27,.01 = " invttail(27,.01)
t_27,.005 = 2.770683  t_27,.01 = 2.4726599

. display "t_27,.025 = " invttail(27,.025) "  t_27,.05 = " invttail(27,.05)
t_27,.025 = 2.0518305  t_27,.05 = 1.7032884
```