Economics 102: Analysis of Economic Data
Cameron Winter 2014 January 30
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**.
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*	2*a*	2*b*	2*c*	2*d*	2*e*	2*f*	3*a*	3*b*	3*c*	3*d*	*Mult.choice*
Points	1	1	1	1	2	1	2	3	1	1	1	3	3	2	2	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 daily data for the difference between the spot price and one-day ahead forward price in the California wholesale electricity market for the one hour period 5-6 p.m. for each day from April 1 1998 to December 31 1998.
`diff` = difference between spot and one-day ahead price in dollars per megawatt hour

(a) Does `diff` appear to be normally distributed? Explain your answer.

(b) What type of diagram would you use to visually display any outlying observations?

(c) What does the following Stata code do?

```
generate y = (diff - diff[_n-1]) / diff[_n-1]
```

(d) The original data were in a called `electricity.dta` Explain how you would read this data into Stata.

(e) If `diff` was actually normally distributed, what range of values would you expect 95% of the observations on `diff` to fall in, given the output on the last sheet?

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 including, for tests, the hypothesis tested and the alternative hypothesis and your conclusion.

(a) Give a **95 percent** confidence interval for population mean price difference.

(b) Give a **90 percent** confidence interval for population mean price difference.

(c) The claim is made that the population mean price difference is zero. Test this claim at significance **level 0.05**.

(d) Give the mathematical formula for the p-value of the test in part (c).

(e) Suppose that a positive price difference one day means that the price difference the next is also likely to be positive. How would this effect your previous analysis? Explain.

(f) Suppose that instead of daily data from April 1998 to December 1998 we only had monthly data. How would this effect your previous analysis? Explain.

3.(a) Consider a simple random sample of size 4 with values $2, 8, 5, 1$.
Compute the sample mean, variance and standard deviation. **Show all workings.**

(b) Let X be the number of company mergers this year in the telecommunications industry. Suppose $X = 0$ with probability 0.4, $X = 1$ with probability 0.2 and $X = 2$ with probability 0.4. Compute the mean, variance and standard deviation of X. **Show all workings.**

(c) Suppose for $X \sim (100, 20^2)$ we form 1000 samples of size 50 and obtain 1000 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?

(d) Consider a random variable that takes values $X = 1$ with probability 0.8 and value $X = 0$ with probability 0.2.
State how you would generate a random sample of size 50. Give a step-by-step method. (You do not need to give exact Stata commands.)

Multiple Choice Questions (1 point each)

1. Data on number of doctor visits in 2012 for a sample of 192 individuals is an example of

a. categorical cross-section data

b. numerical cross-section data

c. categorical time series data

d. numerical time series data

e. none of the above

2. The symmetry statistic is approximately

a. s^3

b. $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3$

c. $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3/s^3$

d. none of the above

3. For data from a simple random sample, $(\bar{x} - \mu)/(s/\sqrt{n})$ is a realization of a random variable that is distributed as

a. t_{n-1} approximately if n is large

b. t_{n-1} if X is normal

c. t_{n-1} always

d. all of a., b., and c.

e. both a. and b.

4. For a simple random sample, $(\bar{X} - \mu)/(\sigma/\sqrt{n})$ is

a. standardized to have mean 0 and variance 1 always

b. normally distributed as $n \to \infty$

c. neither of the above

d. both of the above

5. A key feature of the approach to statistical inference from simple random samples is that

a. the population mean is unchanging but the sample mean changes

b. the sample mean is unchanging but the population mean changes

c. both the population mean and the sample mean change

d. both the population mean and sample mean are unchanging

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* $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$

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

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

Bivariate Data

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

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

$\text{TSS} = \sum_{i=1}^{n}(y_i - \bar{y}_i)^2$ $\quad \text{ErrorSS}=\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2$ $\quad \text{RegSS} = \text{TSS - ErrorSS}$

$R^2 = 1-\text{ErrorSS/TSS}$

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

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

$y|x = x^* \in b_1 + b_2x^* \pm t_{\alpha/2;n-2} \times s_e \times \sqrt{\frac{1}{n} + \frac{(x^*-\bar{x})^2}{\sum_i (x_i-\bar{x})^2} + 1}$

$\text{E}[y|x = x^*] \in b_1 + b_2x^* \pm t_{\alpha/2;n-2} \times s_e \times \sqrt{\frac{1}{n} + \frac{(x^*-\bar{x})^2}{\sum_i (x_i-\bar{x})^2}}$

Multivariate Data

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

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

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

$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$ *and* $F = \frac{(SSE_r - SSE_u)/(k-g)}{SSE_u/(n-k)}$

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

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

```
. sum diff, detail

                            diff
-------------------------------------------------------------
      Percentiles      Smallest
 1%    -100.7092      -107.1325
 5%     -36.7719        -104.92
10%     -20.9986      -100.7092       Obs                 274
25%    -8.998913      -61.88319       Sum of Wgt.         274

50%      -2.2891                      Mean           1.421385
                        Largest       Std. Dev.      35.51708
75%     2.870001       162.9905
90%      19.2143         175.58       Variance       1261.463
95%     70.69516            180       Skewness       2.505035
99%       175.58        215.399       Kurtosis       14.55209

. mean diff

Mean estimation                     Number of obs    =     274

--------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
        diff |   1.421385   2.145666     -2.802769    5.645539
--------------------------------------------------------------

. display "t_273,.005 = " invttail(273,.005) "  t_273,.01 = " invttail(273,.01)
t_273,.005 = 2.5939578  t_273,.01 = 2.3400846

. display "t_273,.025 = " invttail(273,.025) "  t_273,.05 = " invttail(273,.05)
t_273,.025 = 1.9686916  t_273,.05 = 1.6504543
```