Economics 102 A01-A04: Analysis of Economic Data
Cameron Fall 2021 October 14
Department of Economics, U.C.-Davis

First Midterm Exam (Version A)

Compulsory. Closed book. Total of 35 points and worth 20% of course grade.
Read question carefully so you answer the question.
You are to use only **simple calculations (+, -, /, *, square root) and show all workings**.
Use the calculators provided by the department.
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	$1a$	$1b$	$1c$	$1d$	$1e$	$1f$	$2a$	$2b$	$2c$	$2d$	$2e$	$3a$	$3b$	$3c$	$3d$	$3e$	*Mult.choice*
Points	1	1	1	2	2	2	1	2	3	1	3	3	2	1	2	3	5

1.(a) Suppose you have data on the outcome of tossing a six-sided die (taking possible values 1, 2, 3, 4, 5 or 6) for many tosses of the tie. Which Stata command provides a better graphical presentation of the data: `kdensity` or `histogram, discrete`?.

(b) For what sort of dataset would you use the Stata command `use`?

(c) What transformation would you use to make right-skewed data more symmetric?

(d) Suppose a dataset on x_i, $i = 1, ..n$, has mean $\bar{x}$ and sample standard deviation s. What is gained by forming the variable $y_i = (x_i - \bar{x})/s$?

(e) Calculate $\sum_{i=1}^{3} \left(1 + \frac{6}{i}\right)$. **Show all workings.**

(f) Consider a simple random sample of size 3 with values $3, 7$ and 8.
Compute the sample mean and the sample variance. **Show all workings.**

QUESTION 2 USES STATA OUTPUT GIVEN AT THE END OF THIS EXAM. In some cases the answer may be given directly in the output. In other cases you will need to use the output plus additional computation.

The variable `spending` measures individual annual spending on outpatient health services (measured in 2011 $) for a sample of U.S. individuals aged 25 to 65 years.

2.(a) Give the lower quartile for variable `spending`.

(b) Draw with justification the likely shape of a smoothed histogram for this sample. **Explain your answer.**

(c) Provide a **90 percent** confidence interval for population mean annual spending.

(d) What Stata command would directly provide a 95 percent confidence interval for the mean annual spending? **Provide the complete command.**

(e) The claim is made that mean annual health spending of 25-65 year-olds is $1,500. Test this claim at **significance level 0.05**.
State clearly the null and alternative hypotheses and your conclusion.

3.(a) State the three assumptions made about the random variable X_i that are made in the course notes and in class to yield the mean and variance of $\bar{X}$.

(b) Suppose $X = 10$ with probability 0.5, $X = 20$ with probability 0.2 and $X = 30$ with probability 0.3. What is the mean and variance of X? **Show all workings.**

(c) You are given the Stata commands

```
set obs 10000
generate u = runiform()
generate x = 0
replace x = 1 if u < 1/6
```

Give the approximate sample mean for variable `u`.

(d) Suppose for the fair coin toss experiment we form 1000 simple random samples of size 50, from these calculate 1000 95% confidence intervals, and find that 60 of these confidence intervals do not include the true population mean $\mu = 0.5$. Is this surprising? **Explain.**

(e) Consider a simple random sample of size 25 where X has mean 200 and variance 100. Give the mean, variance and standard deviation of $\overline{X}$.

Multiple Choice Questions (1 point each)

1. Data on sales this quarter by each of 23 sales representatives is an example of

a. panel data

b. time series data

c. cross-section data.

2. For a simple random sample of variable X with mean μ and variance σ^2, the standard error of the mean is

a. an estimate of the standard deviation of X

b. an estimate of the standard deviation of $\bar{X}$

c. an estimate of the standard deviation of μ

d. none of the above.

3. Suppose X is normally distributed with mean 100 and variance 25. Then in a large random sample we expect around 95% of observations to lie in the range

a. (75, 125)

b. (50, 150)

c. (95, 105)

d. (90, 110)

e. none of the above.

4. Under simple random sampling, and regardless of sample size, the variable $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$

a. by construction has mean 0 and variance 1

b. by construction has a normal distribution

c. neither of the above

d. both of the above.

5. For the 1880 Census example studied in class and in notes

a. individual age is approximately normally distributed

b. average age in repeated samples of size 25 is approximately normally distributed

c. both of the above

d. neither of the above.

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* $t = \dfrac{\bar{x} - \mu^*}{s/\sqrt{n}}$

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

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

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

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

$R^2 = 1-$ ResidualSS/TSS

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

$$t = \frac{b_2 - \beta_2}{s_{b_2}} \qquad s_{b_2}^2 = \frac{s_e^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad s_e^2 = \tfrac{1}{n-2}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2$$

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

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

Multivariate Data

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

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

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

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

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

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

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

```
. summarize spending, detail

         Outpatient medical spending in 2011 dollars
-------------------------------------------------------------

      Percentiles      Smallest
 1%            0              0
 5%            0              0
10%            0              0       Obs               1,065
25%      252.725              0       Sum of wgt.       1,065

50%      803.975                      Mean           1656.005
                        Largest       Std. dev.      2488.079
75%     1965.036       13965.52
90%     4394.667       14301.15       Variance        6190538
95%     6517.702       15616.38       Skewness       4.278705
99%     12307.95       35034.65       Kurtosis       38.47848

. di _n "Square Root of " _N " = " sqrt(_N) _n   ///
>       "Square Root of " _N-1 " = " sqrt(_N-1) _n

Square Root of 1065 = 32.634338
Square Root of 1064 = 32.619013


. di _n "KEY CRITICAL VALUES FOR THIS EXAM" _n _n       ///
>       "t_" dof ",.005 = " %5.3f invttail(dof,.005) _n ///
>       "t_" dof ",.01  = " %5.3f invttail(dof,.01) _n  ///
>       "t_" dof ",.025 = " %5.3f invttail(dof,.025) _n ///
>       "t_" dof ",.05  = " %5.3f invttail(dof,.05) _n  ///
>       "t_" dof ",.10  = " %5.3f invttail(dof,.10) _n

KEY CRITICAL VALUES FOR THIS EXAM

t_1064,.005 = 2.580
t_1064,.01  = 2.330
t_1064,.025 = 1.962
t_1064,.05  = 1.646
t_1064,.10  = 1.282
```