Economics 102A: Analysis of Economic Data
Cameron Winter 2019 January 31
Department of Economics, U.C.-Davis

First Midterm Exam (Version A)

Compulsory. Closed book. Total of 35 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**.
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	$1a$	$1b$	$1c$	$1d$	$1e$	$1f$	$1g$	$2a$	$2b$	$2c$	$2d$	$2e$	$3a$	$3b$	$3c$	$3d$	$3e$	$Mult.choice$
Points	1	1	1	1	1	2	2	1	1	3	1	3	3	2	2	2	3	5

1.(a) Suppose you have data on income for many individuals. Which Stata command provides a better graphical presentation of the data: `histogram` or `kdensity`?.

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

(c) 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 $y_i = (x_i - \bar{x})/s$?

(d) For what sort of data is a pie chart most useful?

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

(f) Calculate $\sum_{i=1}^{3} \frac{36}{i^2}$. **Show all workings.**

(g) Consider a simple random sample of size 4 with values $12, 15, 12$ and 13.
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 `income` measures the monthly household income (in South African Rand) for a sample of South African households.

2.(a) Give the Stata command that would give additional statistics such as the median for variable `income`.

(b) For these data which is likely to be higher - the mean or the median? **Explain your answer.**

(c) Provide a **90 percent** confidence interval for the mean monthly household income.

(d) What Stata command would directly provide a 95 percent confidence interval for the mean monthly household income?

(e) The claim is made that mean monthly household income in South Africa is 1500. 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 0.1, $X = 20$ with probability 0.8 and $X = 30$ with probability 0.1. 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 < 0.1
```

Draw the likely histogram for variable `x`.

(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 940 of these confidence intervals include the true population mean $\mu = 0.5$. Is this surprising? **Explain.**

(e) Consider a random sample of size 100 where X has mean 20 and variance 400. What approximate range of values do you expect $\bar{x}$ to fall in with 95% probability**?**

Multiple Choice Questions (1 point each)

1. In general it is easiest to establish cause and effect with

- **a.** experimental data
- **b.** observational data
- **c.** neither are adequate for establishing cause and effect
- **d.** both are just as useful for establishing cause and effect.

2. The standard error of the mean obtained from a simple random sample is

- **a.** $\sigma/\sqrt{n}$
- **b.** σ
- **c.** $s/\sqrt{n}$
- **d.** s

3. Desirable properties of an estimator are

- **a.** consistency
- **b.** large variance
- **c.** both of the above
- **d.** neither of the above.

4. For the t statistic based on the sample mean $\bar{X}$ to be exactly $T(n-1)$ distributed in small samples it is needs to be assumed that

- **a.** X is $T(n)$ distributed
- **b.** X is $T(n-1)$ distributed
- **c.** X is normally distributed
- **d.** none of the above.

5. Suppose in a large sample we find that a t-test statistic takes value 1.9. Then the p-value for a two-sided test is

- **a.** between 0.75 and 1.0
- **b.** between 0.5 and 0.75
- **c.** between 0.25 and 0.5
- **d.** between 0.0 and 0.25.

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

Univariate Data

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

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

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

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

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

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

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

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

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

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

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

Multivariate Data

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

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

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

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

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

$\text{Ftail}(df1, df2, f) = \Pr[F > f]$ 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 income

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      income |        840    1428.673    1494.044   34.79167   9510.479

. di _n "Square Root of 840 = " sqrt(840) _n   ///
>       "Square Root of 839 = " sqrt(839) _n

Square Root of 840 = 28.982753
Square Root of 839 = 28.965497


. 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_839,.005 = 2.582
t_839,.01  = 2.331
t_839,.025 = 1.963
t_839,.05  = 1.647
t_839,.10  = 1.283
```