Econ 102_A (Analysis of Economic Data): Cameron Winter 2019
Solutions to First Midterm Exam

Version A

1.(a) `kdensity` is best (as it produces a smoothed histogram and data here are continuous).

(b) For delimited datasets (with comma or space) such as .csv files created by a spreadsheet.

(c) This creates a standardized z-score with sample mean 0 and sample standard deviation 1.

(d) Categorical data (with a few categories)

(e) Natural logarithm (or just logarithm).

(f) $\sum_{i=1}^{3} 36/i^2 = 36/1 + 36/4 + 36/9 = 36 + 9 + 4 = 49.$

(g) $\bar{x} = (12 + 15 + 12 + 13)/4 = 13$
$s^2 = \frac{1}{3}\{(12-13)^2 + (15-13)^2 + (12-13)^2 + (13-13)^2\} = \frac{1}{3}\{1 + 4 + 1 + 0\} = 6/3 = 2.$

2.(a) `summarize income, detail`

(b) The `mean` as income data are typically right skewed.

(c) A 90 percent confidence interval is $1428.67 \pm t_{.05;839} \times 1494.04/\sqrt{840} =$
$= 1428.67 \pm 1.647 \times 51.549 = 1428.67 \pm 84.90 = (1343.77, 1513.56).$

(d) `mean income`

(e) $H_0 : \mu = 1500$ against $H_0 : \mu \neq 1500$ at level 0.05.
$t = (1428.67 - 1500)/(1494.04/\sqrt{840}) = -71.33/51.549 = -1.384.$
Since $|t| = 1.384 < t_{.025;839} = 1.963$ do not reject H_0.
We do not reject $H_0 : \mu = 1500$ at significance level .05.

3.(a) One point each. (1) Common mean $\mathrm{E}[X_i] = \mu$; (2) Common variance $\mathrm{Var}[X_i] = \sigma^2$; (3) Independence of the X_i.

(b) $\mu = \mathrm{E}[X] = 0.1 \times 10 + 0.8 \times 20 + 0.1 \times 30 = 10 + 16 + 3 = 20.$
$\sigma^2 = \mathrm{E}[(X-\mu)^2] = 0.1 \times (10-20)^2 + 0.8 \times (20-20)^2 + 0.1 \times (30-20)^2$
$= 0.1 \times 100 + 0.8 \times 0 + 0.1 \times 100 = 10 + 0 + 10 = 20.$

(c) The histogram for `x` will have two spikes - a larger one at 0 and a smaller one at 1.

(d) Not surprising. For a 95% confidence interval we expect roughly 95% of 1000 or 950 to include μ. 940 is quite close.

(f) $\bar{X}$ has mean 20, variance $400/100 = 4$ and standard deviation 2. And $\bar{X}$ is approximately normally distributed (by the central limit theorem). So in the range $20 \pm 2 \times 2 = (16, 24)$.

Multiple Choice for Version A

Question	**1.**	**2.**	**3.**	**4.**	**5.**
Answer Version A	*a*	*c*	*a*	*c*	*d*

The course grade will be based on a curve from the combined scores of midterm 1 (22.5%), midterm 2 (22.5%), final (45%) and assignments (10%). **The curve for this exam is only a guide**. Suggested average GPA for this course is 2.7. Curve below has average GPA 2.68 for this exam.

Scores out of	35
75*th* percentile	29 (83%)
Median	26.5 (76%)
25*th* percentile	22.5 (64%)

A+	35 and above	C+	24 and above
A	30.5 and above	C	22.5and above
A-	28.5 and above	C-	21.5 and above
B+	27.5 and above	D+	20.5 and above
B	26 and above	D	19 and above
B-	25 and above	D-	18 and above

Version B of 102A

1.(a) $\bar{x} = (12 + 18 + 15)/3 = 15$
$s^2 = \frac{1}{2}\{(12-15)^2 + (18-15)^2 + (15-15)^2\} = \frac{1}{2}\{9 + 9 + 0\} = 18/2 = 9.$

(b) This creates a standardized z-score with sample mean 0 and sample standard deviation 1.

(c) Categorical data (with a few categories).

(d) For delimited datasets (with comma or space) such as .csv files created by a spreadsheet.

(e) `kdensity` is best (as it produces a smoothed histogram and data here are continuous).

(f) Natural logarithm (or just logarithm).

(g) $\sum_{i=1}^{3}(1 + 2i^3) = (1 + 2 \times 1^3) + (1 + 2 \times 2^3) + (1 + 2 \times 3^3) = 3 + 17 + 55 = 75.$

2.(a) `summarize income, detail`

(b) The `median` as income data are typically right skewed.

(c) A 90 percent confidence interval is $1375.22 \pm t_{.05;624} \times 1352.67/\sqrt{625} =$
$= 1375.22 \pm 1.647 \times 54.107 = 1375.22 \pm 89.11 = (1286.11, 1464.33).$

(d) `mean income`

(e) $H_0 : \mu = 1200$ against $H_0 : \mu \neq 1200$ at level 0.05.
$t = (1200 - 1375.22)/(1352.67/\sqrt{625}) = -175.22/54.107 = 3.238.$
Since $|t| = 3.238 > t_{.025;624} = 1.964$ reject H_0.
We reject $H_0 : \mu = 1200$ at significance level .05.

3.(a) The histogram will have just two spikes - a lower one at 0 and a higher one at 1.

(b) $\bar{X}$ has mean 30, variance $900/100 = 9$ and standard deviation 3. And $\bar{X}$ is approximately normally distributed (by the central limit theorem). So in the range $30 \pm 2 \times 3 = (24, 36)$.

(c) $\mu = \mathrm{E}[X] = 0.1 \times 1 + 0.8 \times 2 + 0.1 \times 3 = 0.1 + 1.6 + 0.3 = 2.$
$\sigma^2 = \mathrm{E}[(X - \mu)^2] = 0.1 \times (1-2)^2 + 0.8 \times (2-2)^2 + 0.1 \times (3-2)^2$
$= 0.1 \times 1 + 0.8 \times 0 + 0.1 \times 1 = 0.1 + 0 + 0.1 = 0.2.$

(d) This is surprising. For a 95% confidence interval we expect roughly 95% of 1000 or 950 to include μ. 740 is quite different.

(e) One point each. (1) Common mean $\mathrm{E}[X_i] = \mu$; (2) Common variance $\mathrm{Var}[X_i] = \sigma^2$; (3) Independence of the X_i.

Multiple Choice for Version B

Question	**1.**	**2.**	**3.**	**4.**	**5.**
Answer Version B	*b*	*c*	*a*	*d*	*d*

The course grade will be based on a curve from the combined scores of midterm 1 (22.5%), midterm 2 (22.5%), final (45%), quizzes (5%) and assignments (5%). **The curve for this exam is only a guide**. Suggested average GPA for this course is 2.7. Curve below has average GPA 2.68 for this exam.

Scores out of	35
75*th* percentile	29 (83%)
Median	26.5 (76%)
25*th* percentile	22.5 (64%)

A+	35 and above	C+	24 and above
A	30.5 and above	C	22.5and above
A-	28.5 and above	C-	21.5 and above
B+	27.5 and above	D+	20.5 and above
B	26 and above	D	19 and above
B-	25 and above	D-	18 and above