Econ 102 (Analysis of Economic Data): Cameron Winter 2014
Department of Economics, U.C.-Davis

Solutions to First Midterm Exam

Version A

1.(a) No. Skewed (skewness = 2.5) and large kurtosis (= 14.5 > 3) and from histogram much more peaked amd long tails compared to normal.
(b) A box-and-whisker plot (or more simply a box plot).
(c) This creates the proportionate change in `diff` from one day to the next.
(d) This is a Stata dataset. In Stata use menu `file` > `open` or command `use`.
(e) Expect 95% within two s.d.'s of mean. So $1.42 \pm 2 \times 35.52 = (-69.62, 72.46)$.

2.(a) From `mean diff` output 95% confidence interval is $(-2.80, 5.64)$.
(b) 90 percent confidence interval is $\bar{x} \pm t_{273,.05} \times s/\sqrt{n} = 1.421 \pm \text{invttail}(273, .05) \times 2.146$ $= 1.421 \pm 1.650 \times 2.146 = 1.421 \pm 3.542 = (-2.12, 4.96)$.
(c) $H_0 : \mu = 0$ versus $H_a : \mu \neq 0$, where μ is the population mean difference.
$t = (1.421 - 0)/(s/\sqrt{n}) = 1.421/2.146 = 0.662$.
We reject H_0 at level .05 if $|t| > t_{273,.025} = \text{invttail}(273, 0.025) = 1.969$.
So do not reject H_0 since $|0.662| < 1.969$.
Conclude that data do not reject the claim that population mean differnce equals 0 at level 0.05.
(d) $p = \Pr[|T_{273}| > 0.662]$.
(e) The sample is no longer a simple random sample, violating our assumptions. (This will lead to the standard error of $\bar{x}$ being greater than $s/\sqrt{n}$).
(f) Now we have only $n = 9$ observations. The standard error will be much larger so the confidence interval will be much larger. Furthermore, the t_8 distribution may proide a poor approximation to the distribution of the t-statistic.

3.(a) $\bar{x} = (2 + 8 + 5 + 1)/4 = 16/4 = 4$.
$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{3}[(-2)^2 + (4)^2 + (1)^2 + (-3)^2] = \frac{1}{3}[4 + 16 + 1 + 9] = 30/3 = 10$.
Sample standard deviation $= \sqrt{10} = 3.16$.
(b) $\mu = \mathrm{E}[X] = 0.4 \times 0 + 0.2 \times 1 + 0.4 \times 2 = 0 + 0.2 + 0.8 = 1$.
$\sigma^2 = \mathrm{E}[(X - \mu)^2] = 0.4 \times (0 - 1)^2 + 0.2 \times (1 - 1)^2 + 0.4 \times (2 - 1)^2 = 0.4 + 0 + 0.4 = 0.8$.
$\sigma = \sqrt{0.8} = 0.89$.
(c) Expect $\bar{x}$ to approximately equal $\mu = 100$.
Expect standard deviation of $\bar{x}$ to approximately equal $\sigma/\sqrt{n} = 20/\sqrt{50} = 20/7.07 = 2.82$.
(d) Set sample size to 50. Generate 50 random uniforms and for each set $X = 1$ if the uniform < 0.8 and $X = 0$ if the uniform > 0.8.
[Stata code (not necessary): `set obs 50` `set seed 10101`
`gen u = runiform()` `gen x = 0` `replace x=1 if u<0.8`]

Multiple Choice for Versions A and B

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

Version B

1.(a) No. Skewed (skewness = 2.5) and large kurtosis (= 12.5 > 3) and from histogram much more peaked amd long tails compared to normal.
(b) A box-and-whisker plot (or more simply a box plot).
(c) This creates the change in `diff` from one day to the next.
(d) This is a coma-separated values file. In Stata use menu `file` > `import` or command `insheet` or command `import`.
(e) Expect 99.7% within three s.d.'s of mean. So $5.02 \pm 3 \times 39.52 = (-113.54, 123.58)$.

2.(a) From `mean diff` output 95% confidence interval is $(0.03, 10.02)$.
(b) 90 percent confidence interval is $\bar{x} \pm t_{242,.05} \times s/\sqrt{n} = 5.022 \pm \text{invttail}(242, .05) \times 2.535$ $= 5.022 \pm 1.651 \times 2.535 = 5.022 \pm 4.185 = (0.84, 9.21)$.
(c) $H_0 : \mu = 0$ versus $H_a : \mu \neq 0$, where μ is the population mean difference.
$t = (5.022 - 0)/(s/\sqrt{n}) = 5.022/2.535 = 1.981$.
We reject H_0 at level .05 if $|t| > t_{242,.025} = \text{invttail}(242, 0.025) = 1.970$.
So reject H_0 since $|1.981| > 1.970$.
(d) $p = \Pr[|T_{242}| > 1.981]$.
Conclude that data reject the claim that population mean differnce equals 0 at level 0.05.
(e) The sample is no longer a simple random sample, violating our assumptions. (This will lead to the standard error of $\bar{x}$ being greater than $s/\sqrt{n}$).
(f) Now we have only $n = 9$ observations. The standard error will be much larger so the confidence interval will be much larger. Furthermore, the t_8 distribution may proide a poor approcimation to the distribution of the t-statistic.

3.(a) $\bar{x} = (2 + 14 + 8 + 0)/4 = 24/4 = 6$.
$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{3}[(-4)^2 + (8)^2 + (2)^2 + (-6)^2] = \frac{1}{3}[16 + 64 + 4 + 36] = 120/3 = 40$.
Sample standard deviation $= \sqrt{40} = 6.32$.
(b) $\mu = \mathrm{E}[X] = 0.4 \times 0 + 0.2 \times 1 + 0.4 \times 2 = 0 + 0.2 + 0.8 = 1$.
$\sigma^2 = \mathrm{E}[(X - \mu)^2] = 0.4 \times (0 - 1)^2 + 0.2 \times (1 - 1)^2 + 0.4 \times (2 - 1)^2 = 0.4 + 0 + 0.4 = 0.8$.
$\sigma = \sqrt{0.8} = 0.89$.
(c) Expect $\bar{x}$ to approximately equal $\mu = 50$.
Expect standard deviation of $\bar{x}$ to approximately equal $\sigma/\sqrt{n} = 10/\sqrt{100} = 10/10 = 1.0$.
(d) Set sample size to 50. Generate 50 random uniforms and for each set $X = 1$ if the uniform $<$ 0.7 and $X = 0$ if the uniform $>$ 0.3.
[Stata code (not necessary): `set obs 50` `set seed 10101`
`gen u = runiform()` `gen x = 0` `replace x=1 if u<0.7`]

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 2.4. Curve below has average GPA 2.34 for this exam.

Scores out of	30	
75th percentile	23	(77%)
Median	20.5	(68%)
25th percentile	18	(62%)

A+	28 and above	C+	20 and above
A	25 and above	C	19 and above
A-	24 and above	C-	18 and above
B+	23 and above	D+	17 and above
B	22 and above	D	16 and above
B-	21 and above	D-	14.5 and above