THE UNIVERSITY OF SYDNEY
SCHOOL OF MATHEMATICS AND STATISTICS

Semester 1	MATH1015 (Biostatistics) QUIZ 2 - PRACTICE PROBLEMS	2013

NOTE: Quiz 2 will be in week 11 (ie. week beginning 21 May) in your tutorial class and worth 15% of your final mark. This is a closed book test but you can access the formulae sheet, binomial and normal tables, *t*-table, basic R commands and other information from the course website. There will be some questions related to use of the statistical package R.
Required knowledge: Material from week 1 to week 10 (both inclusive). However, more weights will be given to the material from week 7 to week 10. A practical knowledge of R is required to answer some questions.
Time allowed: 40 mins.
Instructions: There are five possible answers to each question and only one of them is correct. Mark the answer to each question which you think is correct, in pen.

PRACTICE QUIZ 2

1. A 95% confidence interval for the population mean is (51.2, 62.4). If a 5% level of significance is used in testing the hypotheses $H_0 : \mu = 50$ against $H_1 : \mu \neq 50$, the appropriate conclusion would be that the data are:

   (a) against H_0 (b) consistent with H_0 (c) against H_1 (d) inconclusive evidence
   (e) not enough information is given

2. A random sample of size n has been selected from a population with standard deviation σ. To test a null hypothesis on the population mean, a t-test is appropriate if the population is:

   (a) normal with σ known (b) normal with σ unknown (c) any distribution with σ known
   (d) any distribution with σ unknown (e) none of the above

3. In constructing an interval estimate for a population mean, a sample of 100 observations was used. The interval estimate was 19.76 ∓ 1.32. Had the sample size been 400 instead of 100, the interval estimate would have been:

   (a) 19.76 ∓ 0.33 (b) 19.76 ∓ 0.66 (c) 19.76 ∓ 2.64 (d) 9.88 ∓ 0.66 (e) 4.94 ∓ 0.33

4. The following data were drawn from a normal population.

   15 4 24 8 16 13 9 15 7 22

   A 90% confidence interval for the population mean is closest to:

   (a) 13.3 ∓ 3.75 (b) 13.3 ∓ 2.82 (c) 13.3 ∓ 3.95 (d) 13.3 ∓ 7.66 (e) 13.3 ∓ 2.80

5. A random sample of size 10 is selected from a normal population. The population standard deviation is *unknown*. If a *two-sided* test on the population mean is applied with $\alpha = 0.05$, the conclusion that the data are consistent with H_0 can be drawn if the test statistic lies:

   (a) inside $(-1.96, 1.96)$ (b) outside $(1.833, \infty)$ (c) inside $(-2.262, 2.262)$
   (d) inside $(-1.813, 1.813)$ (e) outside $(-2.228, 2.228)$

6. Which of the following is true about Type I error?

   (a) It is more likely to happen when $\alpha = 0.01$ rather than $\alpha = 0.05$.
   (b) It equals to one minus the level of significance.
   (c) It equals to the power of the test.
   (d) It occurs when the conclusion of sufficient evidence against H_0 is drawn but H_0 is true.
   (e) It occurs when the conclusion of consistency with H_0 is drawn but H_0 is false.

PTO for Q7 to Q15

7. Samples of size 2 are drawn from a population with the following probability mass function:

x	3	5
$p(x)$	0.6	0.4

The sampling distribution of the sample mean $\overline{x}$ is:

(a)

$\overline{x}$	3	5
$p(\overline{x})$	0.36	0.16

(b)

$\overline{x}$	3	4	5
$p(\overline{x})$	0.36	0.24	0.16

(c)

$\overline{x}$	3	4	5
$p(\overline{x})$	0.36	0.48	0.16

(d)

$\overline{x}$	3	5
$p(\overline{x})$	0.16	0.36

(e)

$\overline{x}$	3	4	5
$p(\overline{x})$	0.16	0.48	0.36

8. Summary statistics for two random samples from two independent normal populations are reported below:

Sample 1: $n_1 = 9$ $\overline{x}_1 = 10$ $s_1 = 5$

Sample 2: $n_2 = 11$ $\overline{x}_2 = 12$ $s_2 = 3$

Estimates of mean and standard error for the sample mean difference $\overline{X}_1 - \overline{X}_2$ are closest to:

(a) -2 and 0.8864 (b) 2 and 0.9354 (c) -2 and 1.9039 (d) -2 and 0.9354 (e) -2 and 1.8041

9. To test if the left-handed gripping strengths is *stronger* than the right-handed gripping strengths, a random sample of 10 left-handed persons is selected and the measurements are shown in the following table:

Person i	1	2	3	4	5	6	7	8	9	10	$\bar{d}$	s_d
Left hand	140	90	125	130	95	121	85	97	131	110		
Right hand	138	87	110	132	96	120	86	90	129	100		
Difference d_i	2	3	15	-2	-1	1	-1	7	2	10	3.60	5.461

The test statistic t is closest to:

(a) 2.0846 (b) 1.9777 (c) 1.2072 (d) 0.4092 (e) 0.3883

10. Refer to the previous question, if the test statistic is $t > 0$, then the P-value would be:

(a) $2P(T_9 > t)$ (b) $P(T_{18} > t)$ (c) $P(T_9 < t)$ (d) $P(T_9 > t)$ (e) $2(1 - P(T_{18} > t))$

11. A sample of size 200 will be taken at random from a population with binary outcomes. Given that the population proportion is 0.60, the approximate probability that the sample proportion will be *greater* than 0.58 is closest to:

(a) 0.2810 (b) 0.7190 (c) 0.5163 (d) 0.5900 (e) 0.7167

12. To test if the acceptance rate of production line 1 is higher than the acceptance rate of line 2, two random samples of 100 products each are selected from the two lines and the numbers of acceptable products are 76 and 68 respectively. The test statistic is closest to:

(a) 0.3984 (b) 0.8909 (c) 2.5198 (d) 0.7869 (e) 1.2599

Refer to the following information to answer questions 13-15 using `R`*:*

The data in the file `be23-ty378.txt` consists of two columns. The first and second column store the weights of eggs (in g) from manufacturers A and B respectively. The symbol 'NA' in the second column denotes missing observations.

Hint: You can read the data in `R` with `x = read.table(file='be23-ty378.txt')` or

`x = read.table(file=url('http://www.maths.usyd.edu.au/math1015/r/be23-ty378.txt'))` To access the numerical data in column `i`, you may use the expression `x[,i]` where `i` can be 1 or 2.

13. To test if the mean weight of eggs from the two manufacturers are *different* using `t.test` with equal variance assumption, i.e. `var.equal = T`, the test statistic is closest to:

(a) -1.8472 (b) -1.6876 (c) -9.0717 (d) 0.8065 (e) -1.6863

14. At 0.05 level of significance, the P-value and conclusion of the test are respectively:

(a) 0.0494; against H_0 (b) 0.0988; consistent with H_0 (c) 0.0988; against H_0
(d) 0.0494; consistent with H_0 (e) 0.0988; against H_1

15. The assumptions for the test are:

I. The 2 populations are normal; II. The 2 variances are unknown;
III. The 2 sample sizes are large; IV. The 2 variances are equal.

(a) I and II (b) I and III (c) I, II and III (d) I, II and IV (e) All

End of the Test.

Solution:

1. (a)
2. (b)
3. (b) me=$1.32 \times \sqrt{100}/\sqrt{400} = 0.66$ while sample mean remains unchanged.
4. (a) $\overline{X} = 13.3$, $s = 6.4644$ and $t_{9,.05} = 1.833$. The 90% CI for μ=$(13.3 - 1.833 \times 6.4644/\sqrt{10},\ 13.3 + 1.833 \times 6.4644/\sqrt{10}) = 13.3 \mp 3.75$.
5. (c)
6. (d)
7. (c)

$\overline{x}$	(3+3)/2=3	(3+5)/2=4	(5+5)/2=5
$p(\overline{x})$	$0.6^2 = 0.36$	$2 \times 0.6 \times 0.4 = 0.48$	$0.4^2 = 0.16$

8. (e) Mean=$10 - 12 = -2$. SE=$4.0139 \times \sqrt{\frac{1}{9} + \frac{1}{11}} = 1.8041$ where $s_p = \sqrt{\frac{8(5^2)+10(3^2)}{8+10}} = 4.0139$.
9. (a) $d_i = 2, 3, 15, -2, -1, 1, -1, 7, 2, 10$. $t_{\text{obs}} = \frac{\bar{d}}{s_d/\sqrt{n}} = \frac{3.60-0}{5.461/\sqrt{10}} = 2.0846$
10. (d)
11. (b) Since n is large, $\hat{p} \sim N(0.6, 0.6(1-0.6)/200)$ by CLT. $P(\hat{p} > 0.58) = P(Z > \frac{0.58-0.6}{\sqrt{0.6\times0.4/200}}) = P(Z > -0.58) = P(Z < 0.58) = 0.7190$
12. (e) $Z = \frac{0.76-0.68}{\sqrt{0.72(1-0.72)(\frac{1}{100}+\frac{1}{100})}} = 1.2599$ where $\hat{p} = (76+68)/200 = 0.72$.
13. (e)

```
> t.test(x[,1],x[,2],var.equal=T,paired=F)

        Two Sample t-test

data:  x[, 1] and x[, 2]
t = -1.6863, df = 44, p-value = 0.09882
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -9.0716538  0.8065022
sample estimates:
mean of x mean of y
 58.45833  62.59091
```

14. (b)
15. (d)