TUTORIAL EXERCISE PACKAGE - 2013

MATH1015 - BIOSTATISTICS

Semester 1	**Tutorial Week 8**	2013

A CI for μ based on a random sample of size n from a normal population with unknown variance is given by $\bar{x} \pm t \times \frac{s}{\sqrt{n}}$, where s is the sample sd and t is the appropriate t-table value.

A CI for p based on a random sample of size n is given by $\hat{p} \pm z \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ where $\hat{p}$ is the sample proportion, and z si the appropriate z-table value.

The test statistics for a one-sample t-test is $t_{obs} = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{n-t}$ with small n, unknown σ^2 and normality.

Tutorial discussion: Q2, Q3, Q7, Q10 and Q11 marked with *

1. **Multiple choice** Which of the following statemeqnts is true?
   I. When the margin of error is small, the confidence level is high.
   II. When the margin of error is small, the confidence level is low.
   III. A confidence interval is a type of point estimate.
   IV. A population mean is an example of a point estimate.
   (a) I. only (b) II. only (c) III. only (d) IV. only (e) None of these

2. **Multiple choice** *A survey of 1000 Californians finds reports that 48% of the surveyed are excited by the opportunity to take a statistics class. The 95% confidence interval for the true proportion of Californians who are excited to take a statistics class is (3 dp):
   (a) 0.48 ± 0.0310 (b) 0.48 ± 0.0161 (c) 0.48 ± 0.000 (d) 0.48 ± 1.316 (e) None of these

3. *Use the 95% CI for p in Q2 to test, at 5% significance level, the hypotheses:
   (a) $H_0 : p = 0.5$ against $H_1 : p \neq 0.5$
   (b) $H_0 : p = 0.4$ against $H_1 : p \neq 0.4$

4. **Multiple choice** Suqppose that we wish to test the hypothesis $H_0 : \mu = 6$ against $H_1 : \mu > 6$. Based on 16 observations from a normal population with unknown σ^2 it is found that the sample mean, $\bar{x} = 7$ and sample sd, $s = 2$ Using the statistic $t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$, the observed t under the null hypothesis is:
   (a) 1.2 (b) 4.0 (c) 2.0 (d) 0.0 (e) none of these

5. Find the P-value for the test in Q4.

6. **Multiple choice** Information in Q4 and Q5 tell us that:
   (a) the P-value is small and we have evidence against H_0.
   (b) the P-value is small and the data are consistent with H_0.
   (c) the P-value is large and we have strong evidence against H_0.
   (d) the P-value is large and the data are consistent with H_0.
   (e) none of the above.

7. A sample of size 22 from a normal population (unknown variance), yields a total of 191.4. Find a point esmimate of μ, the population mean. Test the hypothesis that the population mean μ is 10 against the alternative that the mean is smaller than 10 given that the sample sd is $s = 2.25$.
Hint: Set up null H_0 and alternative H_1 hypotheses; evaluate a suitable test statistic under H_0; find the P-value; decide whether this P-value is large or small; draw your conclusion.

8. *An inventor has developed a new, energy-efficient lawn mower engine. He claims that the engine will run continuously for 5 hours (300 minutes) on a single littre of regular petrol. Suppose a simple random sample of 46 engines is tested. The engines run for an average of 295 minutes, with a standard deviation of 20 minutes. Test the null hypothesis that the mean run time is 300 minutes against the alternative hypothesis that the mean run time is not 300 minutes. Use a 0.05 level of significance. (Assume that run times for the population of engines are normally distributed.)

Use `R` to answer Q9 to Q11

9. Find the exact P-value in Q5, 7, 8 using `R` and following the steps given below.

10. *Read the germination data from the file `germination.txt` in R using

```
x = scan(file=url("http://www.maths.usyd.edu.au/math1015/r/germination.txt"))
```

or

Test if the mean is different from 86 at 95% significance level using R. **Hint:** use `t.test(x, mu=86)`.

11. *The `hospital` data of discharged patients contains the following columns:

```
Column  Label
1       ID no.
2       Duration of hospital stay
3       Age
4       Sex  1=male  2=female
5       First temperature following admission
6       First WBC(x1000) following admission
7       Received antibiotic 1=yes  2=no
8       Received bacterial culture 1=yes  2=no
9       Service 1=med 2=surg.
```

Read the data using

```
dat=read.table(file=url("http://www.maths.usyd.edu.au/math1015/r/hospital.txt"),skip=1)
```

Set `age=dat[,3]` to be vectors of `age`.

(a) Find the sample size. Explain why t-test is used in testing certain mean of the data.
(b) Plot a boxplot/histograms of `age` to check whether `age` comes from a symmetric distribution.
(c) Assuming that the distribution of `age` is normal, test whether the average `age` of discharged patient is higher than 40.

`R` Hints for Q9-11

Q9: As we have no available data, we will use the `pt` function. Make sure to specify the correct `df`, and don't forget to subtract from 1 and/or multiply by 2.

Q10-11: In this case we have a set of data, so we will use the `t.test` function. Make sure to select the correct values for the `alternative` and `mu` parameters. The P-value will be part of the output.

1. Book P.134, Q.1-2; 2. Book P.134, Q.5;
3. Book P.135, Q.7; 4. Book P.135, Q.8.