THE UNIVERSITY OF SYDNEY

MATH1905 STATISTICS ADVANCED

Semester 2 **Tutorial Week 8** 2012

1. $X_1, X_2, \ldots, X_{25}$ represents a random sample from a distribution with mean $\mu = 10$ and standard deviation $\sigma = 20$. Indicate which of the following distributions is a good approximation to the distribution of $\bar{X} = \frac{1}{25}\sum_i X_i$.

(a) $\mathcal{N}(10, 20^2)$. (b) standard normal. (c) $\mathcal{N}(10, 0.8)$.

(d) $\mathcal{N}(10, 4)$. (e) $\mathcal{N}\left(10, \frac{20^2}{25}\right)$.

2. X is binomial with $n = 100$ and $p = 0.4$ Which of the following normal distributions is a good approximation to the distribution of X?

(a) $\mathcal{N}(100, 0.4)$. (b) standard normal. (c) $\mathcal{N}(40, 0.16)$.

(d) $\mathcal{N}(40, 24)$. (e) $\mathcal{N}(40, 0.24)$.

3. (Illustration of CLT using R). Type `par(mfrow=c(2,2))` to set your graphics with 4 windows.

(1) Generate 2 samples (i.e. 2 vectors `x` and `y`) of size 100 from an exponential distribution with parameter $\lambda = 1$. E.g. `x=rexp(100,1)` produces 1 sample of size 100. Obtain boxplots and histograms of `x` and `y`. Comment on the shape of the exponential distribution.

(2) Show that the expected value for the exponential is 1 and the variance is also 1.

(3) Use a loop (as below) to generate 80 samples of size 100 from an exponential distribution with parameter $\lambda = 1$. For each samples compute the observed sample mean and store the results in a vector called `sample.mean.obs` (as below)

```
sample.mean.obs=numeric(0)       # a place to store the result
for (i in 1:80){                 # begin for-loop
  x.obs=rexp(100,1)              # observe 100 exp(1)
  sample.mean.obs[i]=mean(x.obs) # compute the sample mean for the ith sample
}                                # end for-loop
```

Type `par(mfrow=c(2,1))` to set your graphic device with 2 windows. Obtain the boxplot and the histogram of `sample.mean.obs`. Comment on the shape of the distribution of the sample mean. Repeat this question with 80 samples of size 1000.

4. A P-value of 0.98 indicates that the null hypothesis is true. Comment.

5. Suppose that the probability of success at each repetition of an experiment is p. Perform a one-sided test of $H_0 : p = 0.6 \quad vs. \quad H_1 : p > 0.6$ given the following data (when using the normal approx. don't forget the continuity correction).

(a) There were 8 successes observed in 10 trials.

(b) There were 40 successes observed in 50 trials.

(c) Comment on your answers.

6. The proportion of deaths due to lung cancer in working males aged 15–64 in Australia during the period 1970–1972 was 10%. There is reason to believe that working for an extended period in a chemical plant increases the risk of lung cancer. Accordingly, several chemical plants were investigated, and it was found that of 90 deaths occurring among 15–64 year old male workers who had worked for at least 1 year in the plants, 19 were due to lung cancer. Report a P-value from this study for testing that the death rate is really 10% against the alternative that the rate is greater than 10%. Interpret your result.

Assignment 1 for MATH1905 STATISTICS (due on Tuesday, 2nd October, in week 9) will consist of selected questions from the Problem Sheets for weeks 1, 2, 3, 4, 5, 6, 7, 8.

1. *From the 1998 Examination.* The proportion of defective items produced by a factory is 0.1. As part of the quality control procedure, a random sample of 12 items is inspected daily.

(a) State the distribution of X, the number of defective items in a random sample of size 12.

(b) Find the probability that there are fewer than two defective items on a particular day.

(c) A new manager introduces work practices which are expected to reduce the proportion of defective items produced. After a settling-in period, he asks for a random sample of 200 items to be inspected for defects. Test the effectiveness of the new work practices if it is found that there are only 11 defective items in the sample.

[Hint: Let p be the proportion of defective items produced after the new work practices are introduced. Set up appropriate null and alternative hypotheses concerning p.]

2. It has been claimed that at least 60% of all purchasers of a certain computer program will call the manufacturer's hotline within one month of purchase. A random sample of 12 purchasers of this software is drawn and 3 of those in the sample had contacted the hotline within one month of purchase. Does this provide evidence that the claim of a 60% contact rate is an overestimate? Let p be the true proportion of all purchasers who contact the hotline.

(a) Set up appropriate hypotheses to perform a statistical test.

(b) Why is a 1-sided test appropriate here?

(c) Calculate an exact P-value based on these data. Interpret your findings.

(d) Show that an approximate P-value is smaller than the exact P-value.

3. The clinically accepted value for mean blood pressure in healthy males aged 18 to 22 years is 120 mm Hg and the accepted standard deviation is 20 mm Hg. Assume that blood pressure for this age group is normally distributed.

(a) What proportion of healthy males of this age have a blood pressure above 145.6 mm Hg?

(b) It is widely believed that examination stress causes blood pressure to rise. To test this theory, 10 healthy male students have their blood pressure taken just prior to a Mathematics examination. The readings are simply recorded as *High* (above 145.6 mm Hg) or *Normal* (below 145.6 mm Hg). The actual measurements are not available.

(i) Set up the appropriate null and alternative hypotheses which might be tested using the data in this incomplete form. Define any symbols you use.

(ii) Test the above claim if 3 of the students are found to have *High* blood pressure.