Statistics 160 Fall 2008 Midterm #2 – Solutions

1. (a) If X denotes the number of teenagers who use instant messaging online and n denotes the number of teenagers sampled, then an estimate of the true proportion of teenagers who use instant messaging online is

$$\tilde{p} = \frac{X+2}{n+4} = \frac{736+2}{981+4} = \frac{738}{985} = 0.7492$$

so that the required 95% confidence interval is

$$\tilde{p} \pm z^* \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}} = \frac{738}{985} \pm 1.96 \sqrt{\frac{\frac{738}{985} \times \frac{247}{985}}{981+4}} = [0.7222, 0.7763].$$

1. (b) If X denotes the number of adults who use instant messaging online and n denotes the number of adults sampled, then an estimate of the true proportion of adults who use instant messaging online is

$$\tilde{p} = \frac{X+2}{n+4} = \frac{511+2}{1217+4} = \frac{513}{1221} = 0.4201$$

so that the required 95% confidence interval is

$$\tilde{p} \pm z^* \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}} = \frac{513}{1221} \pm 1.96 \sqrt{\frac{\frac{513}{1221} \times \frac{708}{1221}}{1217+4}} = [0.3925, 0.4478].$$

1. (c) Since the intervals from **(a)** and **(b)** do not overlap, we conclude that there is sufficient evidence that teenagers are much more likely to use IM online than adults. (In other words, we are 95% confident that the true proportion of adults is no greater that 44.78% whereas we are 95% confident that the true proportion of teenagers is no less than 72.22%.)

2. (a) If μ_1 denotes the true number of hours of housework that women do, and if μ_2 denotes the true mean number of hours of housework that men do, then a 98% confidence interval for the true difference, $\mu_1 - \mu_2$, is given by

$$(\overline{x}_1 - \overline{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where $s_1 = 0.52$, $s_2 = 0.5$, and $t^* = 2.374$ since the degrees of freedom are $df = \min\{n_1 - 1, n_2 - 1\} = \min\{91, 80\} = 80$. Thus, the required 98% CI is

$$(3.6 - 3.2) \pm 2.374 \sqrt{\frac{0.52^2}{92} + \frac{0.5^2}{81}} = [0.2157, 0.5843].$$

(Note that if you used μ_1 for the men and μ_2 for the women, then your resulting confidence interval would be $[-0.5843, -0.2157]$.)

2. (b) A 98% confidence interval means that if we were to repeat the experiment of sampling the number of hours of housework done by 81 men and 92 women and constructing the resulting confidence interval a large number of times, then approximately 98% of those intervals would cover the true mean difference. That is, we are constructing our confidence interval using a method that produces correct results 98% of the time.

2. (c) By the confidence interval–hypotheses test duality, the question being asked is equivalent to testing

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{against} \quad H_a : \mu_1 - \mu_2 > 0.$$

Since this is a one-sided test, our 98% confidence interval puts area 1% in each tail. Thus, since 0 does not fall within our CI, this tells us that there is significance evidence at the $\alpha = 0.01$ level to reject H_0 in favour of H_a. (Note that if you used μ_1 for the men and μ_2 for the women, then the resulting significance level would still be 1%, except that you would be rejecting H_0 in favour of $H_a : \mu_1 - \mu_2 < 0$.)

3. Let μ denote the true *age difference* between a married male and his wife. The null hypothesis is therefore $H_0 : \mu = 0$ and the alternative is $H_a : \mu > 0$. This situation can be analyzed with a one-sample t-test. A simple random sample of 40 married couples may be conducted by Statistics Canada using census data, and the sample mean age difference computed as well as the sample standard deviation age difference. The experimental design is based on a matched-pairs design since the ages of married couples are likely to be correlated. (That is, it is unreasonable to assume that an age difference of 1 year is just as likely as an age difference of 30 years.) In order to use the t-test with a sample of size $n = 40$, it must be the case that there are no extreme outliers. Assuming that Statistics Canada is able to obtain a truly random sample, this is a reasonably adequate design.

4. **(C)** Bob's margin of error is

$$\frac{30}{\sqrt{n_b}}$$

and Carol's margin of error is

$$\frac{30}{\sqrt{n_c}}.$$

Since Carol's margin of error is only half as big as Bob's margin of error, we have

$$\frac{30}{\sqrt{n_c}} = \frac{1}{2} \times \frac{30}{\sqrt{n_b}}.$$

Solving gives $\sqrt{n_c} = 2\sqrt{n_b}$ so that

$$n_c = 4n_b.$$

5. **(C)** Since the standard deviations are roughly equal, in order for the standard errors to be the same, it must be the case that the same *number* of middle-aged male Nova Scotians are included as the *number* of middle-aged male Ontarians. Since 0.1% of the population of Nova Scotia is vastly smaller than 0.1% of the population of Ontario, we must choose a lot smaller percentage of Ontarians in order for the sample sizes to be the same.

6. **(D)** The interpretation of confidence intervals *never* allows us to say anything about the particular numbers, in this case $[33, 45]$, that are observed. The 95% refers to a method that produces a result that covers μ in 95% of all random samples.

7. (a) There are $70 \times 30 \times 80 = 168\ 000$ letters in Hamlet. In order for the monkey to correctly type Hamlet, he must get every letter correct. Since he has a $\frac{1}{26}$ chance of getting a single letter correct, the probability that the monkey types Hamlet is

$$\left(\frac{1}{26}\right)^{168\ 000}.$$

Alternatively, we could recognize this as a binomial distribution with probability $p = \frac{1}{26}$ of success on each trial and $n = 168\ 000$ trials. Thus, if X denotes the number of letters correct, then the monkey types Hamlet if and only if all letters are correct. That is,

$$P(X = 168\ 000) = \binom{168\ 000}{168\ 000}\left(\frac{1}{26}\right)^{168\ 000}\left(\frac{25}{26}\right)^{168\ 000 - 168\ 000} = \left(\frac{1}{26}\right)^{168\ 000}.$$

7. (b) The probability of randomly choosing a particular nucleon in the sun is $1/(1.2 \times 10^{57})$. Since

$$\frac{1}{1.2 \times 10^{57}} < \left(\frac{1}{26}\right)^{168\ 000},$$

choosing a particular nucleon in the sun is much more likely.

7. (c) The parameter that 0.032 estimates is *the proportion of vowels in Hamlet*. (Alternatively, the parameter that 0.032 estimates is *the proportion of vowels in the English language*.) If the monkey were truly typing at random, then we would expect to see about $\frac{5}{26} \approx 0.1923$ as the proportion of pure vowels typed. Since 0.032 is significantly smaller than 0.1923, it appears that the monkey is not a very random typist.

8. The correct answers are: **(A)**, **(D)**, **(E)**.

9. The correct answers are: **(B)**, **(C)**, **(E)**.

10. The correct answers are: **(A)**, **(B)**, **(C)**.

11. The correct answers are: **(B)**, **(E)**.

12. The correct answer is: **(C)**.

13. The correct answers are: **(A)**, **(B)**, **(C)**, **(E)**.

14. Let x denote the weight of a sampled pizza so that $\overline{x}$ denotes the average weight of the 4 sampled pizzas. Since $\sigma = 25$ is known, Michael is performing a z-test so that his test statistic is

$$z = \frac{\overline{x} - \mu}{\sigma/\sqrt{n}} = \frac{\overline{x} - 200}{25/\sqrt{4}}.$$

Michael will reject H_0 if $\overline{x} < 180$ which is equivalent to rejecting H_0 if

$$z = \frac{\overline{x} - 200}{25/\sqrt{4}} < \frac{180 - 200}{25/\sqrt{4}} = -1.60.$$

From Table A we find the corresponding one-tailed area to be 0.0548; hence, Michael is performing his test at the $\alpha = 0.0548$ significance level.

The distribution of scores by problem number are as follows. The score is in parentheses.

- Problem 1: 6(8), 6(7), 15(6), 4(5), 2(4), 1(2), 4(0)
- Problem 2: 2(8), 3(7), 4(6), 6(5), 10(4), 4(3), 7(2), 2(1)
- Problem 3: 5(9), 2(8), 4(6), 4(5), 5(4), 9(3), 6(2), 1(1), 2(0)
- Problem 4: 23(2), 15(0)
- Problem 5: 22(2), 16(0)
- Problem 6: 12(2), 26(0)
- Problem 7: 1(8), 2(7), 8(6), 6(5), 13(4), 5(3), 2(2), 1(0)
- Problem 8: 9(1.5), 23(1), 6(0)
- Problem 9: 5(2), 17(1.5), 14(1), 2(0)
- Problem 10: 4(2), 20(1.5), 13(1), 1(0)
- Problem 11: 9(2), 11(1.5), 15(1), 3(0)
- Problem 12: 14(2), 17(1), 7(0)
- Problem 13: 3(2), 11(1.5), 12(1), 11(0.5), 1(0)
- Problem 14: 4(8), 5(7), 1(6), 6(5), 9(4), 6(2), 2(1), 5(0)