UNIVERSITY OF TORONTO
Faculty of Arts and Science

STA130H1 (Introduction to Statistical Reasoning)
MIDTERM TEST
February 24, 2016, 2:10 p.m.

Duration: 100 minutes. Total points: 44.

**** SOLUTIONS ****

1. Suppose X is a random quantity which equals 0 with probability 1/2, or equals 2 with probability 1/3, or equals 8 with probability 1/6.

   (a) [2] Compute the expected value $E(X)$.

   Solution: *Here* $E(X) = \sum_x x\,P(X = x) = 0 \times (1/2) + 2 \times (1/3) + 8 \times (1/6) = 0 + 2/3 + 4/3 = 2$.

   (b) [2] Compute the variance $Var(X)$.

   Solution: *Here* $Var(X) = \sum_x (x-2)^2\,P(X = x) = (0-2)^2 \times (1/2) + (2-2)^2 \times (1/3) + (8-2)^2 \times (1/6) = 4/2 + 0 + 36/6 = 2 + 6 = 8$.

   (c) [1] Compute the standard deviation $sd(X)$.

   Solution: *Here* $sd(X) = \sqrt{Var(X)} = \sqrt{8} \doteq 2.828$.

2. Suppose Y is a random quantity having normal probabilities with mean 40 and variance 25.

   (a) [2] Compute $P(Y < 44)$. [Hint: don't forget the standard normal probability table included at the end of this test.]

   Solution: *Here* $sd(Y) = \sqrt{Var(Y)} = \sqrt{25} = 5$, *so* $Z = (Y - 40)/5$ *has the standard normal distribution, so* $P(Y < 44) = P((Y - 40)/5 < (44 - 40)/5) = P(Z < 0.8) \doteq 0.7881$ *from the table.*

   (b) [2] Compute $P(Y > 38)$.

   Solution: *Here* $P(Y > 38) = P((Y - 40)/5 > (38 - 40)/5) = P(Z > -0.4) = P(Z < +0.4) \doteq 0.6554$ *from the table.*

3. [3] Suppose we roll an ordinary fair six-sided die. Let X be three times the observed die value plus four. (For example, if the die shows 5, then $X = 3 \times 5 + 4 = 19$.) Compute (with explanation) the expected value $E(X)$.

   Solution: *We know that the expected value of the observed value is 3.5. So, if we multiply the value by 3 and add 4, then the expected value is affected the same way,* *so the new expected value is* $3 \times 3.5 + 4 = 14.5$. *Or, alternatively, we can compute this expected value directly as* $\sum_x x\, P(X = x) = (3 \times 1 + 4)(1/6) + (3 \times 2 + 4)(1/6) + (3 \times 3 + 4)(1/6) + (3 \times 4 + 4)(1/6) + (3 \times 5 + 4)(1/6) + (3 \times 6 + 4)(1/6) = 14.5$.

4. [4] In three or four complete English sentences, without using any technical symbols or equations, explain the basic idea of what a P-value is and what it is for, in simple terms that could be understood by someone who has never taken a statistics course.

   Solution: *You could write, for example: "A P-value arises when we want to test the truth of a certain null hypothesis. The P-value is the probability that we would have obtained a result as extreme as, or more extreme than, the observed result. If the P-value is sufficiently small (e.g. less than 0.05), then we reject the null hypothesis; otherwise, we do not reject it."*

5. A recent poll[1] repoted in the media[2] asked about "happiness". They surveyed $n = 1,530$ Canadian adults, and found that 79% of them reported being happy[3].

   (a) [1] Based on the above, how many of the surveyed adults reported being happy?

   Solution: *The number who reported being happy is* $1,530 \times (79\%) \doteq 1,209$.

   (b) [3] Let p be the true fraction of all Canadian adults who would report being happy, and let $\hat{p}$ be the sample fraction from a survey of this size. Then in terms of p and n, what are the mean and variance and sd of $\hat{p}$?

   Solution: *From class, we know that* $\hat{p}$ *has mean* p*, and variance* $p(1-p)/n = p(1-p)/1,530$*, and sd* $\sqrt{\text{variance}} = \sqrt{p(1-p)/n} = \sqrt{p(1-p)/1,530}$.

6. Consider the happiness poll from the previous question.

   (a) [4] Using the conservative option, compute a 95% confidence interval for p based on the poll's findings.

   Solution: *We know from class that the 95% confidence interval for* p *is from* $\hat{p} - 1.96\, sd$ *to* $\hat{p} + 1.96\, sd$*, i.e. from* $\hat{p} - 1.96\sqrt{p(1-p)/n}$ *to* $\hat{p} + 1.96\sqrt{p(1-p)/n}$*. However, the true fraction* p *is unknown. The conservative option replaces* p *by* $1/2$*, and obtains the estimate* $sd \approx \sqrt{(1/2)(1-1/2)/n} = \sqrt{1/4/1,530} \doteq 0.0128$*. is from* $\hat{p} - 1.96\, 0.0128$ *to* $\hat{p} + 1.96\, 0.0128$*, i.e. from* $0.79 - 1.96\, 0.0128$ *to* $0.79 + 1.96\, 0.0128$*, i.e. from about 0.765 to 0.815, i.e. the interval [0.765, 0.815]. i.e. the interval [76.5%, 81.5%]. (Alternatively, under the conservative option the "95% margin of error" is* $98\%/\sqrt{n} = 98\%/\sqrt{1,530} \doteq 2.5\%$*, so the confidence interval is from about* $79\% - 2.5\%$ *to* $79\% + 2.5\%$*, i.e. from 76.5% to 81.5%.)*

[1] http://angusreid.org/wp-content/uploads/2016/01/2016.01.26-life-satisfaction.pdf
[2] http://www.thestar.com/life/2016/02/01/two-thirds-of-canadians-pretty-happy-poll-finds.html
[3] For "Happy", we combined the responses "Very happy" and "Pretty happy" together.

(b) [2] State your conclusion from the confidence interval as a complete English sentence.

Solution: *We conclude that the true fraction of Canadian adults who would report being happy is probability somewhere between 76.5% and 81.5%.*

7. Consider the happiness poll from the previous two questions.

(a) [2] Consider testing the null hypothesis that $p = 0.77$ against the alternative hypothesis that $p > 0.77$. Specify in a complete English sentence what the P-value would correspond to in this case.

Solution: *Here the P-value would be the probability, under the null hypothesis that the true fraction p is equal to 0.77, that we would observe a fraction $\hat{p}$ which is equal to 0.79 or larger.*

(b) [3] Compute this P-value, using (with explanation) a normal approximation.

Solution: *Since n is large, according to the Central Limit Theorem, the value of $\hat{p}$ is approximately normal, with mean p and sd $\sqrt{p(1-p)/n}$. Under the null hypothesis that $p = 0.77$, the mean is $p = 0.77$, and the sd is $\sqrt{p(1-p)/n} = \sqrt{0.77(1-0.77)/1,530} \doteq 0.0108$.*
So, the P-value is the probability that a normal with mean 0.77 and sd 0.0108 would be 0.79 or larger. This is the same as the probability that its Z-score would be $(0.79 - 0.77)/0.0108$ or larger, i.e. $P(Z > (0.79 - 0.77)/0.0108) \doteq P(Z > 1.85) = 1 - P(Z < 1.85) \doteq 1 - 0.9678 = 0.0322 = 3.22\%$, using the standard normal probabilities table.

(c) [2] Determine (with explanation) whether or not the null hypothesis should be rejected in this case, according to standard scientific practice.

Solution: *Since the P-value is less than 0.05, yes we should reject the null hypothesis.*

(d) [2] State in a complete English sentence your conclusion from this hypothesis test.

Solution: *We conclude that the true fraction of Canadian adults who would report being happy is probably more than 77%.*

8. The above happiness poll also included separate results for each province. In Ontario, they surveyed $n_1 = 511$ adults, and found that 80% of them reported being happy. In Quebec, they surveyed $n_2 = 360$ adults, and found that 77% of them reported being happy. Write p_1 and $\hat{p}_1$ for the true and sample fractions who report being happy in Ontario, and p_2 and $\hat{p}_2$ for Quebec.

(a) [3] In terms of p_1 and p_2 and n_1 and n_2, what are the mean and variance and sd of the sample difference $\hat{p}_2 - \hat{p}_1$?

Solution: *Here the sample difference* $\hat{p}_2 - \hat{p}_1$ *has mean* $p_2 - p_1$, *and variance* $p_1(1-p_1)/n_1+p_2(1-p_2)/n_2$, *and sd* $\sqrt{\text{variance}} = \sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}$.

(b) [4] Using the bold option, compute a 95% confidence interval for the difference $p_2 - p_1$ based on the poll's findings.

Solution: *We know from class that the 95% confidence interval for* $p_2 - p_1$ *is from* $\hat{p}_2 - \hat{p}_1 - 1.96\,sd$ *to* $\hat{p}_2 - \hat{p}_1 + 1.96\,sd$, *i.e. from* $\hat{p} - 1.96\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}$ *to* $\hat{p} + 1.96\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}$. *However, the true fractions* p_1 *and* p_2 *are unknown. For the bold option, we approximate the sd by replacing* p_1 *by* $\hat{p}_1 = 0.80$, *and* p_2 *by* $\hat{p}_2 = 0.77$, *to obtain that* $sd \approx \sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2} = \sqrt{0.80(1-0.80)/511 + 0.77(1-0.77)/360} \doteq 0.0284$. *Then, the 95% confidence interval for the difference* $p_2 - p_1$ *is from* $\hat{p}_2 - \hat{p}_1 - 1.96\,sd$ *to* $\hat{p}_2 - \hat{p}_1 + 1.96\,sd$, *i.e. from* $0.80 - 0.77 - 1.96 \times 0.0284$ *to* $0.80 - 0.77 + 1.96 \times 0.0284$, *i.e. from about* -0.0257 *to* 0.0857, *i.e. the interval* $[-0.0257,\ 0.0857]$, *i.e. the interval* $[-2.57\%,\ 8.57\%]$.

(c) [2] State your final conclusion as a complete English sentence.

Solution: *We conclude that the difference between the true fraction of adults in Ontario who would report being happy, minus the true fraction of adults in Quebec who would report being happy, is probability somewhere between −2.57% and +8.57%. (Or, alternatively: We conclude that the true fraction of adults who would report being happy in Ontario is probably somewhere between 2.57% less than that of Quebec, and 8.57% more than that of Quebec.)*