Solutions for Practice for FINAL EXAM

Math 102 / Core 143

A. Robertson

1. We have three boxes with unknown numbers in each. From each box we draw, with replacement, 1000 times. From Box I we want to investigate the sum; from Box II we want to investigate the average; from Box III we want to investigate the product. We can use the Normal curve to approximate the results for

A) Box I only

B) Box I and II only

C) Box I and III only

D) All three boxes

B: Only the sum of draws (and hence the average) are approximately Normal. The Central Limit Theorem says nothing about products.

2. A student performs a hypothesis test comparing 115 samples against an established norm. She gets a test statistics of 2.14. Which of the following is correct at a significance level of 5%?

A) The p-value is 1.6% so we conclude that our sample is different from the established norm

B) The p-value is 1.6% so we conclude that our sample is *not* different from the established norm

C) The p-value is 2.5% so we conclude that our sample is different from the established norm

D) The p-value is 2.5% so we conclude that our sample is *not* different from the established norm

A: We use the Normal curve to find the area to the right of 2.14 and get 1.6%. Since the p-value is less than 5%, we conclude that the difference is real, and unlikely to be due to chance.

3. Roll a fair 8-sided die 64 times. Find the 84^{th} percentile for the number of threes rolled.

A) 8.00

B) 9.25

C) 10.65

D) 12.80

C: The box is a 0-1 box with 1/8th ones and 7/8ths zeros. We draw 64 times. We have $EV = 8$ and $SE_{sum} = \sqrt{64}\sqrt{(1/8)(7/8)} \approx 2.65$. We want the 84th percentile, which on the Normal chart correspronds to $z = 1.00$ (we want the area between $-z$ and z to be 64% so that we have 84% to the left of z. Converting $z = 1.00$ out of standard units, we get $1.00 = \frac{X-8}{2.65}$ which gives 10.65.

4. You have in your possession a biased coin which has probability of landing heads 1/3.

i) Find the chance that at least one head appears on five consecutive tosses.

A) 32.9%

B) 50.0%

C) 66.7%

D) 86.8%

D: We have a small number of tosses so we can't use the Normal curve. Instead, we use the NOT rule. We don't want all tosses to be tails. The chance of this occurring is $(2/3)^5 \approx .132$ or 13.2%. Hence, the probability we want is $100\% - 13.2\% = 86.8\%$.

ii) Find the chance of getting exactly 8 heads out of 10 tosses.

A) 99.7%

B) 30.0%

C) 5.2%

D) 0.3%

D: We use the Binomial formula with $n = 10$, $k = 8$, and $p = 1/3$. We have $\binom{10}{8}(1/3)^8(2/3)^2 \approx$ .003 or .3%.

5. A remarkable (ficticious) study result in Norway found a correlation between a person's IQ and their height. The average IQ of the study group was 100 with a standard deviation of 15, while the average height of the study group was 66 inches, with a standard deviation of 6 inches. The remarkable results was that the researchers found a correlation coefficient of $r = .3$.

i) What do you predict for the height of a person in the study group with a 110 IQ?

A) 79.3 in

B) 72.0 in

C) 68.9 in

D) 67.2 in

D: Let x be IQ and y be height. We want to predict y given x, so we use:

$$y - a_y = r(SD_y/SD_x)(x - a_x).$$

Hence, we have

$$y - 66 = .3(6/15)(110 - 100),$$

which gives a predicted height of 67.2 in.

ii) What do you predict for the IQ of a person in the study group who is 6 feet tall?

A) 98.6

B) 104.5

C) 126.2

D) 150.0

B: Using the same variables as in part (i), we want to predict x given y, so we use:

$$x - a_x = r(SD_x/SD_y)(y - a_y).$$

Hence, we have

$$x - 100 = .3(15/6)(72 - 66),$$

which gives a predicted IQ of 104.5.

6. A gambling game is based on a simplified roulette wheel. This wheel has two red slots, two black slots, and one green slot. You bet \$1 on red. If red shows you get back \$2, otherwise you lose your \$1. Assuming each slot is equally likely, what is the approximate chance of not losing any money on 100 spins of the wheel?

A) 2%

B) 5%

C) 10%

D) 25%

A: Our distribution is: $1, 1, -1, -1, -1$. We are drawing 100 times. Hence, $EV_{sum} = n \cdot \text{avg} = 100 \cdot (-1/5) = -20$ and $SE_{sum} = \sqrt{100}\sqrt{\frac{2\cdot 1^2 + 3\cdot(-1)^2}{5} - (-1/5)^2} \approx 9.8$. Want the probability that the sum is ≥ 0. Converting $-.5$ to standard units gives $\frac{-.5-(-20)}{9.8} \approx 1.99$. Consulting the Normal table, the area to the right of 2.04 is approximately 2%.

7. Consider the box: $\underline{|0, 1|}$. We are going to draw 9 times, with replacement, and look at the product of the draws (multiply all numbers drawn together). The chance that the product is equal to 1 is

A) 0.2%

B) 1.1%

C) 97.9%

D) 100%

A: To get a product of 1, all draws must be 1. The probability of this happening is $(1/2)^9$ or about .2%.

8. We want to test if the amount of Carbon Monoxide (CO) in the air (in parts per million) is higher that 70. We take 5 readings and get: $78, 83, 68, 72, 88$. At a significance level of 1%, would you conclude that the amount of CO in the air is higher than 70ppm?

Since our sample size is 5, we use a t-test to test:

H_0: Amount is 70ppm
H_A: Amount is $>$ 70ppm

We have an observed average of 77.8 and an expected average, assuming the null hypothesis, of 70. We have $SD^+ = \sqrt{\frac{5}{4}} \cdot \sqrt{\frac{78^2+83^2+68^2+72^2+88^2}{5} - (77.8)^2} \approx 8.07$. Hence, $SE^+_{avg} = \frac{8.07}{\sqrt{5}} \approx 3.61$. This gives our test statistic:

$$t_4 = \frac{77.8 - 70}{3.61} \approx 2.16,$$

where we have $5 - 1 = 4$ degrees of freedom. Consulting the t-table, we have a p-value of between 2.5% and 5%. Since this is not less than 1%, we do not reject H_0. Our conclusion: the amount of CO is not higher than 70ppm at a significance level of 1%.

9. In a small village of 1000, 200 people are surveyed to see if they watch the tv show "Survivor." From this survey it was found that 160 people watch "Survivor." Find a 95% confidence interval for the percentage of people in the village that watch "Survivor."

A 95% CI is $EV_{\%} \pm 1.96 SE_{\%}$. We have $EV_{\%} = \frac{160}{200} \cdot 100\% = 80\%$. We have $SE_{\%} = \sqrt{\frac{1000-200}{1000-1}} \cdot \frac{\sqrt{(.8)(.2)}}{\sqrt{200}} \cdot 100\% \approx 2.53\%$. Hence, the desired CI is $80\% \pm 5.06\%$.

10. A student has a set of data that she believes to be correlated. However, the student (correctly) calculated the correlation coefficient to be $r = .013$ and she concludes that the data is not correlated in any way. Do you agree with her conclusion? State your reasoning. Also, if the correlation coefficient were based on a sample with size 102, would you conclude, at a significance level of 5%, that $r = 0$?

This shows there is little to no *linear* correlation. There could be nonlinear correlation.

As for the hypothesis test, we are testing $H_0 : r = 0$ against $H_a : r > 0$. We have $t_{100} = \frac{.013\sqrt{100}}{\sqrt{1-(.013)^2}} \approx .13$. Since our degrees of freedom is > 25, we use the Normal curve approximation to the t-curve. We clearly have a p-value $> 5\%$. Hence, we would conclude that $r = 0$ by accepting the null hypothesis.

11. We tally some categorical data according to gender and political affiliation. We have the following table:

	Democrat	Republican	Other
Female	104	79	12
Male	113	128	4

Based on this data, find the following probabilities.

i) Probability that a person is female, given that the person is Republican. $79/207 \approx .38$

ii) Probability that a person is a Democrat, given that the person is male. $113/245 \approx .46$

iii) Probability that a person is "Other." $16/440 \approx .036$

12. A recent tv commercial claims that a higher percentage of students get A's when they used the new "Hookd On Fonix" than those students that did not use it. To test this claim we take a sample of 30 students who have used "Hookd On Fonix" and a sample of 50 student who have not used "Hookd On Fonix." We found:

Of those that used "Hookd On Fonix," 18 received A's and 12 did not.

Of those that have not used "Hookd On Fonix," 25 received A's and 25 did not.

At a significance level of 5%, do you agree with the commercial's claim?

This is a 2-sample z-test. We have an observed difference of 10%. For the samples, we have $SE_1 = \frac{\sqrt{(18/30)(12/30)}}{\sqrt{30}} \cdot 100\% \approx 8.94\%$ and $SE_2 = \frac{\sqrt{(1/2)(1/2)}}{\sqrt{50}} \cdot 100\% \approx 7.07\%$. This gives us $SE_{diff} = \sqrt{(8.94)^2 + (7.07)^2} \approx 11.4\%$. Testing H_0: difference $= 0$, versus H_A: difference > 0, we have

$$z = \frac{10\% - 0\%}{11.4\%} \approx .88.$$

This gives a p-value (using the Normal curve) of about 19%, so we don't reject H_0 and contradict the commercial claim.

13. Consider a fair 4 sided die (with sides 1,2,3, and 4). If we roll this die 121 times, what is the approximate probability that we roll at least thirty-five 2s?

This is an application of the Central Limit Theorem. Our distribution is binary with the fraction of 2s equal to 1/4.

Using this, we have *for the sum*: $EV = (1/4)(121) = 30.25$ and $SE = \sqrt{121}\sqrt{(1/4)(3/4)} \approx 4.76$.

We want ≥ 35 twos. Hence, we want the area to the right of 34.5 under the Normal curve with mean 30.25 and SD 4.76. We convert 34.5 to standard units: $\frac{34.5-30.25}{4.76} \approx .89$. So, we want the area to the right of .89 on the standard Normal curve. This is 18.76%.

Note: The final exam will be 25 multiple choice questions based on all material covered in the class.