BMCC Mat 150 Introduction to Statistics OER Version
Sample Final Exam
with solutions and references to the OpenStax "Introduction to Statistics" textbook

These materials were written by Professor Chris McCarthy (BMCC CUNY) for the "Achieving the Dream 2018" Open Educational Resource project. They are free to be used provided attribution is given.

This sample final exam is separated into two parts. Each covering roughly 1/2 a semester's worth of material. Each part can serve as quiz. Together, they can serve as a final exam. The questions are similar to questions that can be found in the OpenStax OER text "Introductory Statistics."

Part I

Covering material from Chapters 1 – 4 in the OpenStax text.

Question 1. This question tests your understanding of statistics terminology.

You are conducting a study to determine the average number of hours BMCC students spend doing homework each day. You decide to gather data for this study by asking the students in your BMCC statistics class how many hours they spend doing homework. In your statistics class there are 15 males and 15 females.

(1.1) In this study the "population" is
(A) students of all ages
(B) college students
(C) BMCC students
(D) the students in your BMCC statistics class.

Answer to 1.1: ______________.

(1.2) In this study the "sample" is
(A) students of all ages
(B) college students
(C) BMCC students
(D) the students in your BMCC statistics class.

Answer to 1.2: ______________.

(1.3) The sampling method used in your study is
(A) biased (because it only includes students taking statistics).
(B) unbiased (because equal number of males and females).

Answer to 1.3: ______________.

(1.4) The number of hours a BMCC student spends doing homework each day is an example of a

(A) quantitative (numerical) variable.
(B) qualitative (categorical) variable.

Answer to 1.4 ______________.

(1.5) The gender of a BMCC student is an example of a

(A) quantitative (numerical) variable.
(B) qualitative (categorical) variable.

Answer to 1.5 ______________.

(1.6) In your study, the mean number of hours the students in your statistics class spend doing homework is an example of a

(A) statistic.
(B) parameter.

Answer to 1.6 ______________.

(1.7) In your study, the mean number of hours the students in BMCC spend doing homework is an example of a

(A) statistic.
(B) parameter.

Answer to 1.7 ______________.

Question 1. Answers and notes.
(1.1) C (1.2) D (1.3) A (1.4) A (1.5) B (1.6) A (1.7) B

Sections 1.1 and 1.2 in the OpenStax text cover the material being tested in Question 1.
For related problems, starting on page 47 of the OpenStax text, see questions 1 – 38.

Question 2. This question is about descriptive statistics and the visual representation of data.

Name of Student	Abe	Ben	Cathy	Dina	Eunice	Fran
Age in years (x)	5	5	8	9	9	12

(2.1) Using the above data, find (a) the mean $\bar{x}$, to the nearest 10^{th}, (b) the median of x, to the nearest 10^{th}, (c) the mode(s) of x, and (d) calculate the sample standard deviation of the children's ages, S_x, to the nearest 100^{th}. Put answers to (a), (b), (c), and (d) on the lines.

(a) $\bar{x} =$ ______________. (b) median = ______________. (c) mode(s) = ______________. (d) $S_x =$ ______________.

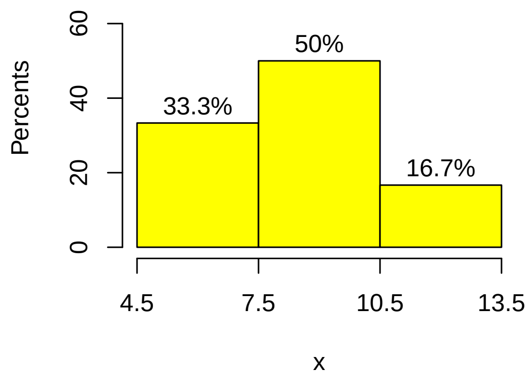
(2.2) Draw the relative frequency histogram for the children's ages x. Use class start = 4.5 years and class width = 3 years. Calculate percentages to nearest 10^{th} of a percent and put percents on top of the bars.

Draw the relative frequency histogram here.

Question 2. Answers and notes.

(2.1) (a) $\bar{x} =$ ___8___. (b) median = ___8.5___. (c) mode = ___bimodal 5, 9___. (d) $S_x =$ ___2.68___.

(2.2) Relative frequency histogram of x.



Sections 2.1, 2.5, 2.7 in the OpenStax text cover the material being tested in Question 2. Note: section 2.3 has material on percentiles.

For related problems, starting on page 127 of the OpenStax text, see questions 12 – 17 (histograms); 42 – 48 (mean, median, mode); 69 – 73 (standard deviation).

Question 3. If 30% of NY college students have taken algebra, what is the probability that in a random sample of 6 NY college students that at least 2 of them have taken algebra? Let X count how many have taken algebra in samples consisting of 6 NY college students. Since the population of NY College students is very large you can assume that this a Bernoulli process. Show work below. Give final answer to nearest 100^{th} of a percent.

Question 3. Answer and notes.

The distribution of X is binomial with $n = 6$ and $p = 30\% = .3$.

$P($ at least 2 have taken algebra $) = P(X \geq 2) = 1 - P(X = 0, 1)$ and $P(X = 0, 1) = P(X = 0) + P(X = 1)$, which we calculate below using the binomial distribution formula:

$$P(X = r) = nCr\ p^r\ (1-p)^{n-r}$$
$$P(X = 0) = 6C0\ (.3)^0\ (.7)^6 = (1)(1)(0.117649) = 0.117649$$
$$P(X = 1) = 6C1\ (.3)^1\ (.7)^5 = (6)(.3)(0.16807) = 0.302526$$

So

$$P(X = 0, 1) = P(X = 0) + P(X = 1) = 0.117649 + 0.302526 = 0.420175$$

and

$$P(\text{ at least 2 have taken algebra }) = P(X \geq 2) = 1 - P(X = 0, 1) = 1 - 0.420175 = 0.579825 \approx \underbrace{57.98\%}_{\text{Answer}}$$

Section 4.3 (Binomial Distribution) in the OpenStax text covers the material being tested in Question 3.

For related problems, starting on page 285 of the OpenStax text, see questions 37 – 44 (binomial distribution).

Part II

Covering material found in Chapters 5 – 12 in the OpenStax text.

Question 4. If X is $N(80, 10)$ find $P(80 < X < 92.5)$. Also, draw X, z, with the relevant regions for finding the probability shaded.

Question 4. Answer and notes.

Probability calculation:

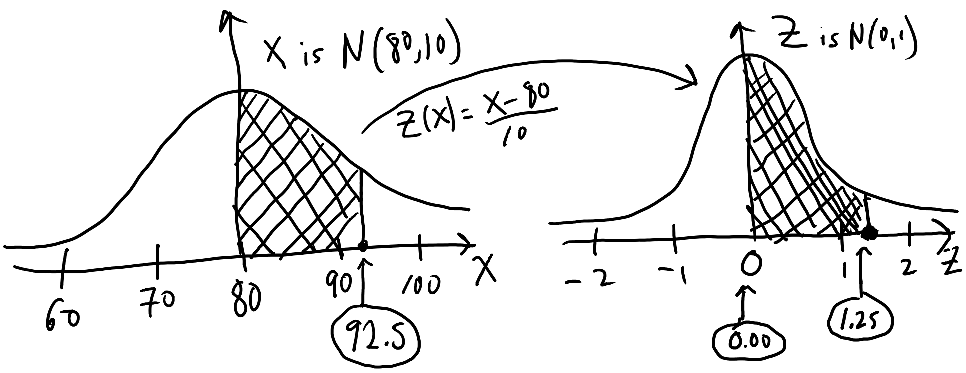
$$\begin{aligned} P(80 < X < 92.5) &= P(z(80) < z(X) < z(92.5)) \\ &= P(0.00 < z < 1.25) \\ &= A(1.25) - A(0.00) \\ &= .8944 - .5000 = .3944 \quad \leftarrow \quad \textit{answer} \end{aligned}$$

z-transformation of X

$$z(X) = \frac{X - \mu_X}{\sigma_X} = \frac{X - 80}{10}$$

$$z(80) = \frac{80 - 80}{10} = \frac{0}{10} = 0.00$$

$$z(92.5) = \frac{92.5 - 80}{10} = \frac{12.5}{10} = 1.25$$



Chapter 6 (Normal Distribution) in the OpenStax text covers the material being tested in Question 4.

For related problems, see the questions at the end of Chapter 6 (starting on page 385 OpenStax). They are all on the normal distribution.

Question 5. From extensive studies, we know that the weights of male giraffes are normally distributed, with a mean of 2500 pounds and a standard deviation of 250 pounds. If we weigh 25 randomly chosen male giraffes, what is the probability the mean weight of the giraffes in our sample will be between 2475 and 2525 pounds? Also, draw $\overline{X}, z$, with the relevant regions for finding the probability shaded.

Question 5. Answer and notes.

Let X(giraffe) = its weight in pounds. We are told in the problem that X is normally distributed, so we immediately know, by the CLT, that $\overline{X}$ is $N(\mu_{\overline{X}}, \sigma_{\overline{X}})$ with $\mu_{\overline{X}} = \mu_X = 2500$ pounds and

$$\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{250}{\sqrt{25}} = \frac{250}{5} = 50 \text{ pounds}$$

So $\overline{X}$ is $N(2500, 50)$. Then:

Probability calculations:

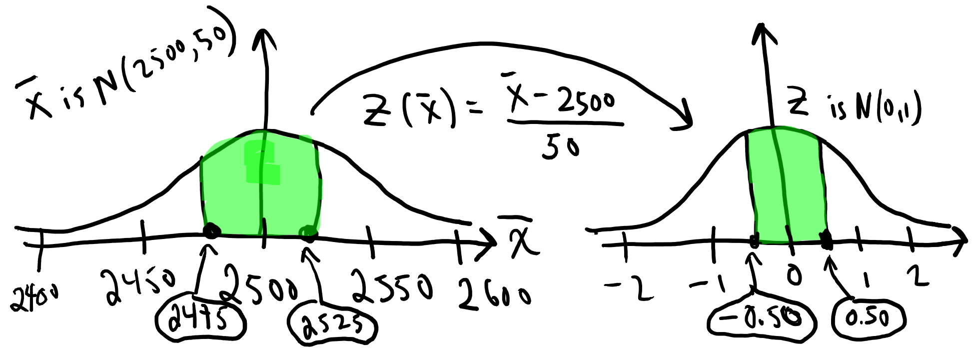
$$\begin{aligned} P(2475 < \overline{X} < 2525) &= P(z(2475) < z(\overline{X}) < z(2525)) \\ &= P(-0.50 < z < 0.50) \\ &= A(0.50) - A(-0.50) \\ &= .6915 - .3085 = .3830 = 38.3\% \quad \leftarrow \quad \textit{answer} \end{aligned}$$

z-transformation of $\overline{X}$

$$z(\overline{X}) = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{\overline{X} - 2500}{50}$$

$$z(2475) = \frac{2475 - 2500}{50} = \frac{-25}{50} = -0.50$$

$$z(2525) = \frac{2525 - 2500}{50} = \frac{25}{50} = 0.50$$



The above figure is of $\overline{X}$ and z.

Section 7.1 (Central Limit Theorem for Sample Means) and 7.3, in the OpenStax text covers the material being tested in Question 5.

For related problems, see questions 62 – 71 (starting on page 430) at the end of Chapter 7 (OpenStax).

Question 6. The Eastern American Toad (*Bufo a. americanus*) is a smallish toad found throughout New England. In a sample of 81 adult female Eastern American Toads the mean weight was $\overline{X} = 43.5\ g$ and the standard deviation was $S_X = 15.1\ g$. Using this data find the 95% confidence interval for the mean weight of adult female Eastern American Toads.

Question 6. Answer and notes.

The only slightly difficult part of this problem is figuring out t_*. To figure out t_* we look in the t-distribution table. To use the t-distribution table you need to know which row to look in and which column:

The row is the degrees of freedom (d.f.), which for finding the confidence interval for the mean (like in this problem) is $n-1$. So the $d.f. = n-1 = 81-1 = 80$. So we look in the row "80". We want the 95% confidence interval, so we look in column that says 95% confidence level. So we get $t_* = 1.990$.

Then we just plug $n = 81$, $\overline{X} = 43.5\ g$; $S_X = 15.1\ g$; and $t_* = 1.990$ into the CI formula: $CI = \overline{X} \pm t_* \frac{S_X}{\sqrt{n}}$, which becomes:

95% CI for $\mu_x = \overline{X} \pm t_* \frac{S_X}{\sqrt{n}} = 43.5 \pm 1.990 \frac{15.1}{\sqrt{81}} = 43.5 \pm 1.990 \frac{15.1}{9} = \underline{43.5\ g\ \pm\ 3.34\ g}$. ← answer.

Section 8.2 (A Single Population Mean using the t distribution (Confidence Intervals)) in the OpenStax text covers the material being tested in Question 6.

For related problems, see questions 104 – 116 (starting on page 485) at the end of OpenStax Chapter 8.

Question 7. The Eastern American Toad (*Bufo a. americanus*) is a smallish toad found throughout New England.

It is claimed that the mean weight of females of this species is greater than 40 g. Test this claim at the $\alpha = .05$ level of significance using the following data.

In a sample of 81 adult female Eastern American Toads the mean weight was $\overline{X} = 43.5\ g$ and the standard deviation was $S_X = 15.1\ g$.

Answer (choose (a) or (b) and fill in the blank lines about the p-value):

(a) The data provides statistically significant evidence that the claim is true (______ < p-value < ______).

(b) The data fails to provides statistically significant evidence that the claim is true (______ < p-value < ______).

Question 7. Answer and notes.

(claim) $H_A : \mu_X > 40\ g$

(null) $H_0 : \mu_X = 40\ g$

The p-value is the maximum probability of making a "type I" error if we are willing to accept the claim as true based upon the statistics from the sample. The maximum probability of getting data that will fool us into making type I error will happen if the claim is false by as small an amount as possible, i.e., if H_0 is true. Finally, for this problem, if we are willing to accept the H_A as true based on $\overline{X}= 43.5$, we should also be willing to accept the claim as true if $\overline{X}\geq 43.5$. So, using the CLT (since $n > 30$) and the t-distribution approximation of standard normal distribution we get:

$$p - value = P(\overline{X}\geq 43.5 \mid \mu_X = 40) = P\left(t \geq t(43.5)\right) = P\left(t \geq 2.0858\right)$$

where

$$t\left(\overline{X}\right) = \frac{\overline{X} - \mu_X}{S_X/\sqrt{n}}$$

and so

$$t(43.5) = \frac{43.5 - 40}{15.1/\sqrt{81}} = 2.0858.$$

The degrees of freedom $df = n - 1 = 81 - 1 = 80$.

Using the t-distribution table we see that on row 80, that 2.0858 falls between 1.990 and 2.374, going to the top of the table (1-tailed), we see that $P(t \geq 1.990) = .025$ and that $P(t \geq 2.374) = .001$. So, the p-value is between .001 and .025. In other words, the $p - value < .05$.

So, we get the answer:

(a) The data provides statistically significant evidence that the claim is true ($\underline{.001} < p - value < \underline{.025}$).

Chapter 9 (Hypothesis Testing) in the OpenStax text covers the material being tested in Question 7.

Question 8. Using linear regression (least squares), find the equation of the line $y = a_0 + a_1 x$ which best fits the following data: $(x, y) = \{(1, 1),\ (2, 4),\ (2, 5),\ (3, 6),\ (3, 7)\}$. Then plot the data and graph the regression line.

Use the formulas:

$$a_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} \qquad a_0 = \frac{\sum y - a_1 \sum x}{n}$$

Question 8. Answer and notes.

$n = 5$ and

$\sum x = 1 + 2 + 2 + 3 + 3 = 11.$

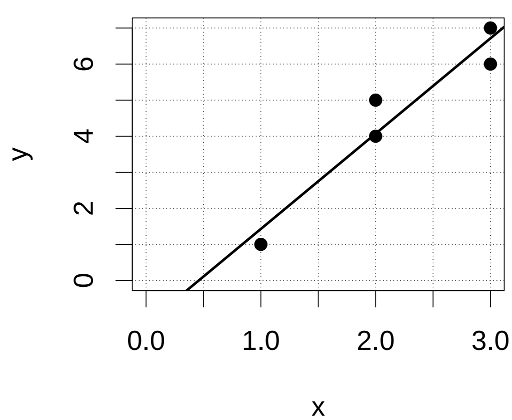
$\sum y = 1 + 4 + 5 + 6 + 7 = 23.$

$\sum xy = 1 + 8 + 10 + 18 + 21 = 58.$

$\sum x^2 = 1^2 + 2^2 + 2^2 + 3^2 + 3^2 = 1 + 4 + 4 + 9 + 9 = 27.$

$$a_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \frac{5(58) - (11)(23)}{5(27) - (11)^2} = 2.642857 \text{ and } a_0 = \frac{\sum y - a_1 \sum x}{n} = \frac{23 - 2.642857(11)}{5} = -1.214285$$

So the regression line's equation $y = a_0 + a_1 x$ is $\qquad y = -1.214285 + 2.642857x \qquad \leftarrow$ *answer*.



Chapter 12 (Regression and Correlation) in the OpenStax text covers the material being tested in Question 8. For related problems, see questions at the end of Chapter 12 OpenStax, starting on page 719.

Question 9. Qualitative Correlation Questions (Chose the best answer.)

(9.1) Suppose that the number of ticks carried by deer in the spring is positively correlated with the average winter temperature. This means:

(a) if the average winter temperature increases, we should expect that the number of ticks carried by deer will increase in the spring.

(b) if the average winter temperature increases, we should expect that the number of ticks carried by deer will decrease in the spring.

(c) knowing if the average winter temperature has increased will not allow us to better predict whether there will be an increase, or decrease, in the number of ticks carried by dear in the spring.

(9.2) Suppose that the number of cases of the flu is negatively correlated with the air temperature. This means:

(a) as the air temperature drops the number of cases of the flu increases.

(b) as the air temperature drops the number of cases of the flu decreases.

(c) knowing the air temperature will not help us to predict how many cases of the flu there will be.

(9.3) Suppose that the number of mosquitoes is uncorrelated with the amount of soda consumed in an area. This means:

(a) the more soda consumed, the more mosquitoes.

(b) the more soda consumed, the fewer mosquitoes.

(c) if the amount of soda consumed changes, we should not expect a change in the number of mosquitoes.

Question 9. Answer and notes.
(9.1) (a)
(9.2) (a)
(9.3) (c)

Chapter 12 (Regression and Correlation) in the OpenStax text covers the material being tested in Question 9. For related problems, see questions at the end of Chapter 12 OpenStax, starting on page 719.
