Math 369 / 650 Fall 2020 Midterm Examination One

Professor Adam Kapelner

Wednesday, September 30, 2020

Code of Academic Integrity

Since the college is an academic community, its fundamental purpose is the pursuit of knowledge. Essential to the success of this educational mission is a commitment to the principles of academic integrity. Every member of the college community is responsible for upholding the highest standards of honesty at all times. Students, as members of the community, are also responsible for adhering to the principles and spirit of the following Code of Academic Integrity.

Activities that have the effect or intention of interfering with education, pursuit of knowledge, or fair evaluation of a student's performance are prohibited. Examples of such activities include but are not limited to the following definitions:

Cheating Using or attempting to use unauthorized assistance, material, or study aids in examinations or other academic work or preventing, or attempting to prevent, another from using authorized assistance, material, or study aids. Example: using an unauthorized cheat sheet in a quiz or exam, altering a graded exam and resubmitting it for a better grade, etc.

By taking this exam, you acknowledge and agree to uphold this Code of Academic Integrity.

Instructions

This exam is 75 minutes (variable time per question) and closed-book. You are allowed **one** page (front and back) of a "cheat sheet", blank scrap paper and a graphing calculator. Please read the questions carefully. No food is allowed, only drinks.

Problem 1 [8min] (and 8min will have elapsed) The following are questions about testing and power.

- [9 pt / 9 pts] Record the letter(s) of all the following that are **true**. At least one will be true.
 - (a) Type II errors are only possible if H_0 is true
 - (b) Type II errors are only possible if H_a is true
 - (c) If you set a higher α , the probability of making a Type II error increases (if H_a is true)
 - (d) If you set a higher α , the probability of making a Type II error decreases (if H_a is true)
 - (e) As n increases, the probability of making a Type I error increases (if H_0 is true)
 - (f) As n increases, the probability of making a Type II error increases (if H_a is true)
 - (g) A lower α setting makes the p-value come out larger.
 - (h) A lower α setting makes null hypothesis rejections more "statistically significant".
 - (i) Imagine you are doing a two-tailed two-sample (or two-proportion) test. Let "effect size" denote $\theta_1 \theta_2$. If you are trying to prove a large effect size, the power is higher than if you trying to prove a small effect size.

Your answer will consist of a string (e.g. aebgd) where the order of the letters does not matter nor does upper / lowercase.

bdhi

Problem 2 [5min] (and 13min will have elapsed) The rats used in most laboratories are called Sprague-Dawley rats (the "lab rat"), a special breed of brown rat because they are calmer and easier to handle. 30 Sprague-Dawley rats are purchased from the company known to be the best national lab rat supplier.

- [7 pt / 16 pts] Record the letter(s) of all the following that are **true**. At least one will be true.
 - (a) The 30 rats is likely a representative sample of all rats if the sampling was done by the supplier via simple random sampling.
 - (b) The 30 rats is likely a representative sample of all brown rats if the sampling was done by the supplier via simple random sampling.
 - (c) The 30 rats is likely a representative sample of all Sprague-Dawley rats if the sampling was done by the supplier via simple random sampling.
 - (d) The 30 rats could be representative of all Sprague-Dawley rats even if the sampling was done by the supplier without simple random sampling.
 - (e) The population size is N = 30.
 - (f) The population size is infinite.
 - (g) There is no definite population, but you assume a population and consider this population to be infinite if you invoke the population sampling assumption.

Your answer will consist of a string (e.g. aebgd) where the order of the letters does not matter nor does upper / lowercase.

cdg

Problem 3 [5min] (and 18min will have elapsed) Same as before. The rats used in most laboratories are called Sprague-Dawley rats (the "lab rat"), a special breed of brown rat because they are calmer and easier to handle. 30 Sprague-Dawley rats are purchased from the rat supplier. Animal studies are frequently done on rats since rats are considered a model for humans. Consider a nutritional study that tests the effect of magnesium supplementation on cardiovascular disease. In this study, the n = 30 rats are given 10mg of magnesium daily. Since the lifespan of rats is on average 2 years, this multi-year study waits until all the rats die until they do the data collection. The data is whether or not each rat had heart problems during their life (i.e. each rat either did have a heart problem or did not have a heart problem).

- [10 pt / 26 pts] Record the letter(s) of all the following that are **true**. At least one will be true.
 - (a) The data collected is commonly denoted x_1, x_2, \ldots, x_{30} .
 - (b) The data collected is commonly denoted X_1, X_2, \ldots, X_{30} .
 - (c) The DGP is most likely $\stackrel{iid}{\sim} \mathcal{N}(\theta, \sigma^2)$ and θ measures how long the rats live.
 - (d) The DGP is most likely $\stackrel{iid}{\sim} \mathcal{N}(\theta, \sigma^2)$ and θ measures the probability of heart problems.
 - (e) The DGP is most likely $\stackrel{iid}{\sim}$ Bernoulli (θ) and θ is mean life length measured in years.
 - (f) The DGP is most likely $\stackrel{iid}{\sim}$ Bernoulli (θ) and θ is the probability of at least one lifetime heart problem.
 - (g) The DGP is most likely hypergeometric and thus the rat measurements are dependent.
 - (h) The DGP is most likely $\stackrel{iid}{\sim}$ with mean 10mg.
 - (i) The DGP is most likely $\stackrel{iid}{\sim}$ with mean 2yr.
 - (j) The researcher's intent is most likely to use the data to make inference about population or DGP parameter(s).



Problem 4 [4min] (and 22min will have elapsed) Same as before. The rats used in most laboratories are called Sprague-Dawley rats (the "lab rat"), a special breed of brown rat because they are calmer and easier to handle. 30 Sprague-Dawley rats are purchased from the rat supplier. Animal studies are frequently done on rats since rats are considered a model for humans. Consider a nutritional study that tests the effect of magnesium supplementation on cardiovascular disease. In this study, the n = 30 rats are given 10mg of magnesium daily. Since the lifespan of rats is on average 2 years, this multi-year study waits until all the rats die until they do the data collection. The data is whether or not each rat had heart problems during their life (i.e. each rat either did have a heart problem or did not have a heart problem). Thus the DGP is $\stackrel{iid}{\sim}$ Bernoulli (θ) and θ is the probability of at least one heart problem and we denote the data x_1, x_2, \ldots, x_{30} .

- [6 pt / 32 pts] Record the letter(s) of all the following that are **true**. At least one will be true.
 - (a) θ is a parameter of the DGP.
 - (b) θ is a realization from a rv.
 - (c) θ is a point estimate.
 - (d) The value of θ is known before the study begins and this study will only confirm it.
 - (e) The value of θ is unknown before the study begins.
 - (f) The value of θ is unknown before the study begins but will be known after the study is over as that is the purpose of this study.

Your answer will consist of a string (e.g. aebgd) where the order of the letters does not matter nor does upper / lowercase.

ae

Problem 5 [6min] (and 28min will have elapsed) Same as before. The rats used in most laboratories are called Sprague-Dawley rats (the "lab rat"), a special breed of brown rat because they are calmer and easier to handle. 30 Sprague-Dawley rats are purchased from the rat supplier. Animal studies are frequently done on rats since rats are considered a model for humans. Consider a nutritional study that tests the effect of magnesium supplementation on cardiovascular disease. In this study, the n=30 rats are given 10mg of magnesium daily. Since the lifespan of rats is on average 2 years, this multi-year study waits until all the rats die until they do the data collection. The data is whether or not each rat had heart problems during their life (i.e. each rat either did have a heart problem or did not have a heart problem). Thus the DGP is $\stackrel{iid}{\sim}$ Bernoulli (θ) and θ is the probability of at least one heart problem and we denote the data x_1, x_2, \ldots, x_{30} .

- [9 pt / 41 pts] Record the letter(s) of all the following that are **true**. At least one will be true.
 - (a) We can use the data to compute a point estimate $\hat{\theta}$ which is the best numeric guess of the value of θ .
 - (b) The point estimate of θ is a realization from the rv denoted X.
 - (c) The point estimate of θ is a realization from the rv denoted $\hat{\theta}$.
 - (d) The point estimate of θ is a realization from the sampling distribution.
 - (e) To compute the point estimate of θ , you need to presuppose an H_a .
 - (f) A reasonable point estimate of θ is the proportion of x_i 's that are equal to one.
 - (g) A reasonable point estimate of θ is the proportion of x_i 's that are equal to zero.
 - (h) A reasonable point estimate of θ is \bar{x} .
 - (i) A reasonable point estimate of θ is $\hat{\sigma}^2$.

Your answer will consist of a string (e.g. aebgd) where the order of the letters does not matter nor does upper / lowercase.

acdfh

Problem 6 [11min] (and 39min will have elapsed) Same as before. The rats used in most laboratories are called Sprague-Dawley rats (the "lab rat"), a special breed of brown rat because they are calmer and easier to handle. 30 Sprague-Dawley rats are purchased from the rat supplier. Animal studies are frequently done on rats since rats are considered a model for humans. Consider a nutritional study that tests the effect of magnesium supplementation on cardiovascular disease. In this study, the n=30 rats are given 10mg of magnesium daily. Since the lifespan of rats is on average 2 years, this multi-year study waits until all the rats die until they do the data collection. The data is whether or not each rat had heart problems during their life (i.e. each rat either did have a heart problem or did not have a heart problem). Thus the DGP is $\stackrel{iid}{\sim}$ Bernoulli (θ) and θ is the probability of at least one heart problem and we denote the data x_1, x_2, \ldots, x_{30} . The point estimate we will use for θ is \bar{x} and the estimator is \bar{X} .

- [12 pt / 53 pts] Record the letter(s) of all the following that are **true**. At least one will be true.
 - (a) The estimator is biased.
 - (b) The estimator is asymptotically unbiased.
 - (c) The estimator is unbiased.
 - (d) The estimator has an MSE of zero for some values of $\theta \in (0, 1)$.
 - (e) The largest MSE of the estimator is 1/(4n).
 - (f) Consider $\ell(\hat{\theta}, \theta) = 0$ if $\hat{\theta} = \theta$ and 1 otherwise. This is a legal loss function.
 - (g) The loss function in (f) is a reasonable loss function that you can use to compare other estimators to \bar{X} .
 - (h) The risk under the loss function in (f) is equal to the variance.
 - (i) Consider $\ell(\hat{\theta}, \theta) = |\hat{\theta} \theta|$. This is a legal loss function.
 - (j) The loss function in (i) is a reasonable loss function that you can use to compare other estimators to \bar{X} .
 - (k) The risk under the loss function in (i) is equal to the variance.
 - (l) The estimator $\hat{\theta} = \frac{1}{2} (\max\{x_1, x_2, \dots, x_{30}\} + \min\{x_1, x_2, \dots, x_{30}\})$ will have similar MSE to \bar{X} .



Problem 7 [4min] (and 43min will have elapsed) Same as before. The rats used in most laboratories are called Sprague-Dawley rats (the "lab rat"), a special breed of brown rat because they are calmer and easier to handle. 30 Sprague-Dawley rats are purchased from the rat supplier. Animal studies are frequently done on rats since rats are considered a model for humans. Consider a nutritional study that tests the effect of magnesium supplementation on cardiovascular disease. In this study, the n = 30 rats are given 10mg of magnesium daily. Since the lifespan of rats is on average 2 years, this multi-year study waits until all the rats die until they do the data collection. The data is whether or not each rat had heart problems during their life (i.e. each rat either did have a heart problem or did not have a heart problem). Thus the DGP is $\stackrel{iid}{\sim}$ Bernoulli (θ) and θ is the probability of at least one heart problem and we denote the data x_1, x_2, \ldots, x_{30} . The point estimate we will use for θ is \bar{x} and the estimator is \bar{X} . We wish to prove that the incidence of heart problems in the rats given magnesium is less than 48% (the national average for heart problems in the American adult population).

- [11 pt / 64 pts] Record the letter(s) of all the following that are **true**. At least one will be true.
 - (a) $H_a: \theta < 0.48$
 - (b) $H_a: \theta > 0.48$
 - (c) $H_a: \theta \neq 0.48$
 - (d) $H_0: \theta \le 0.48$
 - (e) $H_0: \theta > 0.48$
 - (f) $H_0: \theta = 0.48$
 - (g) $\alpha = 5\%$ is the scientific community's standard.
 - (h) $\alpha = 2.5\%$ in the left tail is the scientific community's standard.
 - (i) A target power of 1 is desirable and achievable.
 - (j) A target power of 1α is desirable and achievable.
 - (k) A target power of α is desirable and achievable.



Problem 8 [7min] (and 50min will have elapsed) Same as before. The rats used in most laboratories are called Sprague-Dawley rats (the "lab rat"), a special breed of brown rat because they are calmer and easier to handle. 30 Sprague-Dawley rats are purchased from the rat supplier. Animal studies are frequently done on rats since rats are considered a model for humans. Consider a nutritional study that tests the effect of magnesium supplementation on cardiovascular disease. In this study, the n = 30 rats are given 10mg of magnesium daily. Since the lifespan of rats is on average 2 years, this multi-year study waits until all the rats die until they do the data collection. The data is whether or not each rat had heart problems during their life (i.e. each rat either did have a heart problem or did not have a heart problem). Thus the DGP is $\stackrel{iid}{\sim}$ Bernoulli (θ) and θ is the probability of at least one heart problem and we denote the data x_1, x_2, \ldots, x_{30} . The point estimate we will use for θ is \bar{x} and the estimator is \bar{x} . We wish to prove that the incidence of heart problems in the rats given magnesium is less than 48%, the national average for heart problems in the American adult population. Thus $H_a: \theta < 0.48$ and $H_0: \theta \geq 0.48$. Upon the study's completion, the researchers compute $\hat{\theta} = 9/30 = 0.3$.

- [7 pt / 71 pts] Record the letter(s) of all the following that are **true**. At least one will be true.
 - (a) The binomial test is an exact test of the hypothesis of interest.

Let $B \sim \text{Binomial}(30, 48\%)$ with PMF $p_B(x)$ and CDF $F_B(x)$. The following is an abridged table of the PMF and CDF. The values are rounded to the nearest two digits but should be treated as exact.

	x	7				11											
	$p_B(x)$																0
_	$F_B(x)$	0	0.01	0.04	0.08	0.14	0.24	0.37	0.52	0.66	0.78	0.87	0.93	0.97	0.99	1	1_

- (b) The scientific standard of $\alpha = 5\%$ is attainable in the binomial test.
- (c) A retainment region of $\{0, 1, \dots, 9\}$ is the tregion that most closely provides the scientific standard of $\alpha = 5\%$.
- (d) A rejection region of $\{0, 1, \dots, 9\}$ is the the region that most closely provides the scientific standard of $\alpha = 5\%$.
- (e) At $\alpha = 0.04$, the test rejects the null hypothesis.
- (f) At $\alpha = 0.04$, the test retains the null hypothesis.
- (g) Fisher's p value is 1%.



Problem 9 [5min] (and 55min will have elapsed) Same as before. The rats used in most laboratories are called Sprague-Dawley rats (the "lab rat"), a special breed of brown rat because they are calmer and easier to handle. 30 Sprague-Dawley rats are purchased from the rat supplier. Animal studies are frequently done on rats since rats are considered a model for humans. Consider a nutritional study that tests the effect of magnesium supplementation on cardiovascular disease. In this study, the n = 30 rats are given 10mg of magnesium daily. Since the lifespan of rats is on average 2 years, this multi-year study waits until all the rats die until they do the data collection. The data is whether or not each rat had heart problems during their life (i.e. each rat either did have a heart problem or did not have a heart problem). Thus the DGP is $\stackrel{iid}{\sim}$ Bernoulli (θ) and θ is the probability of at least one heart problem and we denote the data x_1, x_2, \ldots, x_{30} . The point estimate we will use for θ is \bar{x} and the estimator is \bar{X} . We wish to prove that the incidence of heart problems in the rats given magnesium is less than 48%, the national average for heart problems in the American adult population. Thus $H_a: \theta < 0.48$ and $H_0: \theta \ge 0.48$. Upon the study's completion, the researchers compute $\hat{\theta} = 9/30 = 0.3$. We wish to test by using the one-proportion z test at $\alpha = 1\%$. Note that $\Phi(-2.33) = 1\%$.

- [8 pt / 79 pts] Record the letter(s) of all the following that are **true**. At least one will be true.
 - (a) $\hat{\theta} \mid H_0 \sim \mathcal{N} (0.48, 0.48(1 0.48))$
 - (b) $\hat{\theta} \mid H_0 \sim \mathcal{N} (0.48, 0.48(1 0.48)/30)$
 - (c) $30(\hat{\theta} \mid H_0 0.48)/(0.48(1 0.48)) \sim \mathcal{N}(0, 1)$
 - (d) $\sqrt{30}(\hat{\theta} \mid H_0 0.48) / \sqrt{0.48(1 0.48)} \sim \mathcal{N}(0, 1)$
 - (e) The retainment region is $\hat{\theta} \geq .27$ (to the nearest two digits)
 - (f) The retainment region is $z \ge -2.33$ on the standardized scale
 - (g) The test rejects the null hypothesis.
 - (h) The test retains the null hypothesis.

Your answer will consist of a string (e.g. aebgd) where the order of the letters does not matter nor does upper / lowercase.

bdefh

Problem 10 [5min] (and 60min will have elapsed) Same as before. The rats used in most laboratories are called Sprague-Dawley rats (the "lab rat"), a special breed of brown rat because they are calmer and easier to handle. 30 Sprague-Dawley rats are purchased from the rat supplier. Animal studies are frequently done on rats since rats are considered a model for humans. Consider a nutritional study that tests the effect of magnesium supplementation on cardiovascular disease. In this study, the n = 30 rats are given 10mg of magnesium daily. Since the lifespan of rats is on average 2 years, this multi-year study waits until all the rats die until they do the data collection. The researchers also were interested in mean life expectancy of the magnesium-supplemented rats. Regardless of what θ was before, we now denote mean life expectancy as θ . The lifespans in years of each rat were 1.09, 2.48, 3.08, 2.57, 1.04, 0.87, 4.18, 2.23, 3.22, 1.33, 2.49, 1.69, 3.18, 1.39, 2.52, 4.8, 2.44, 1.47, 2.64, 3.96, 3.08, 2.71, 2.8, 3.4, 3.86, 2.28, 3.65, 3.28, 1.54, 1.94. Here are two statistics: $\bar{x} = 2.57$ and s = 1.00. We wish to test if these rats lived longer than the average life expectancy of 2 years.

- [9 pt / 88 pts] Record the letter(s) of all the following that are **true**. At least one will be true.
 - (a) $H_a: \theta > 2$
 - (b) $H_a: \theta \neq 2$
 - (c) We can use the one sample z test to run this test without any assumptions.
 - (d) We can use the one sample z test to run this test by assuming an $\stackrel{iid}{\sim} \mathcal{N}(\theta, 1^2)$ DGP.
 - (e) We can use the one sample z test to run this test by assuming an $\stackrel{iid}{\sim} \mathcal{N}(\theta, \sigma^2)$ DGP if σ^2 were given to you.
 - (f) We can use the one sample z test to run this test by assuming an $\stackrel{iid}{\sim} \mathcal{N}(\theta, \sigma^2)$ DGP where σ^2 is an unknown constant.
 - (g) We can use the one sample t test to run this test by assuming an $\stackrel{iid}{\sim} \mathcal{N}(\theta, 1^2)$ DGP.
 - (h) We can use the one sample t test to run this test by assuming an $\stackrel{iid}{\sim} \mathcal{N}(\theta, \sigma^2)$ DGP if σ^2 were given to you.
 - (i) We can use the one sample t test to run this test by assuming an $\stackrel{iid}{\sim} \mathcal{N}(\theta, \sigma^2)$ DGP where σ^2 is an unknown constant.

Your answer will consist of a string (e.g. aebgd) where the order of the letters does not matter nor does upper / lowercase.

aei

But the question text had an error when the exam was given so this wasn't counted in the midterm grade.

Problem 11 [5min] (and 65min will have elapsed) Same as before. The rats used in most laboratories are called Sprague-Dawley rats (the "lab rat"), a special breed of brown rat because they are calmer and easier to handle. 30 Sprague-Dawley rats are purchased from the rat supplier. Animal studies are frequently done on rats since rats are considered a model for humans. Consider a nutritional study that tests the effect of magnesium supplementation on cardiovascular disease. In this study, the n=30 rats are given 10mg of magnesium daily. Since the lifespan of rats is on average 2 years, this multi-year study waits until all the rats die until they do the data collection. The researchers also were interested in mean life expectancy of the magnesium-supplemented rats. Regardless of what θ was before, we now denote mean life expectancy as θ . The lifespans in years of each rat were 1.09, 2.48, 3.08, 2.57, 1.04, 0.87, 4.18, 2.23, 3.22, 1.33, 2.49, 1.69, 3.18, 1.39, 2.52, 4.8, 2.44, 1.47, 2.64, 3.96, 3.08, 2.71, 2.8, 3.4, 3.86, 2.28, 3.65, 3.28, 1.54, 1.94. Here are two statistics: $\bar{x}=2.57$ and s=1.00. We wish to test if these rats lived longer than the average life expectancy of 2 years. Hence $H_a: \theta > 2$ and $H_0: \theta \le 2$. We will use the one-sample t-test and use $\alpha=1\%$. Note that $F_{T_{29}}(-2.46)=1\%$.

- [6 pt / 94 pts] Record the letter(s) of all the following that are **true**. At least one will be true.
 - (a) The rejection region is $\hat{\theta} > 4.46$ (to the nearest two digits)
 - (b) The rejection region is $\hat{\theta} > 5.03$ (to the nearest two digits)
 - (c) The rejection region is $\hat{\theta} > 2.45$ (to the nearest two digits)
 - (d) The rejection region is $\hat{\theta} > 3.02$ (to the nearest two digits)
 - (e) You can conclude from this test that the power in this test is very high (i.e. near 1).
 - (f) The one-sample t-test is an approximate test.

Your answer will consist of a string (e.g. aebgd) where the order of the letters does not matter nor does upper / lowercase.

С

Problem 12 [10min] (and 75min will have elapsed) Same as before. [But no space to put the old text] Researchers repeated this study with a higher dose of magnesium on 6 rats. The lifespan of these "higher" dose rats where 3.11, 2.38, 3.34, 0.85, 3.65, 3.89. Here are their statistics: $\bar{x} = 2.87$ and s = 1.11. The statistics for the previous sample of rats (who received the 10mg dose which we now called the "lower" dose) was n = 30, $\bar{x} = 2.57$ and s = 1.00. We want to test if there's any difference in rat lifespan between the two doses. We will assume the same DGP (iid normal) for the higher dose group (but with a different mean than the lower dose group). We will also assume the same variance for both groups. Note that $F_{T_{34}}(-2.44) = 1\%$ and $F_{T_{6,70}}(-3.04) = 1\%$.

- [11 pt / 105 pts] Record the letter(s) of all the following that are **true**. At least one will be true.
 - (a) The standard error of the sampling distribution of $\hat{\theta}_{higher} \hat{\theta}_{lower}$ is 1.04 (to the nearest two digits)
 - (b) The standard error of the sampling distribution of $\hat{\theta}_{higher} \hat{\theta}_{lower}$ is 1.02 (to the nearest two digits)
 - (c) The standard error of the sampling distribution of $\hat{\theta}_{\text{higher}} \hat{\theta}_{\text{lower}}$ is 0.46 (to the nearest two digits)
 - (d) The standard error of the sampling distribution of $\hat{\theta}_{higher} \hat{\theta}_{lower}$ is 0.20 (to the nearest two digits)
 - (e) The standard error of the sampling distribution of $\hat{\theta}_{higher} \hat{\theta}_{lower}$ is 0.24 (to the nearest two digits)
 - (f) The standard error of the sampling distribution of $\hat{\theta}_{higher} \hat{\theta}_{lower}$ is 0.49 (to the nearest two digits)
 - (g) The standard error of the sampling distribution of $\hat{\theta}_{higher} \hat{\theta}_{lower}$ is 0.22 (to the nearest two digits)
 - (h) You were given sufficient information to compute an exact rejection region
 - (i) You were given sufficient information to compute a retainment region at $\alpha = 1\%$
 - (j) You were given sufficient information to compute a retainment region at $\alpha = 2\%$
 - (k) You were given sufficient information to compute a retainment region at $\alpha = 5\%$

Your answer will consist of a string (e.g. aebgd) where the order of the letters does not matter nor does upper / lowercase.

hj(c is deleted since it's subject to rounding error)