Practice Exam Solutions

Friday, April 3, 2020 8:31 AM

Practice - Midterm...

Midterm II: Ch. 3.4-3.5, 4.1, 4.3, 5.1-5.3, 6.1 STAT 140 Practice Exam - SOLUTIONS

Name:

Instructions:

- Write your name and section on this cover page.
- Turn off your cell phone and put it away.
- You **may** use a calculator. However, you **may not** use a calculator on your phone or any other device that connects to the internet.
- You must show all your work. Purely numerical answers with no notation and no steps shown will not receive credit.
- You have **50 minutes** to complete the exam.
- You are expected to obey the Honor Code while taking this test. You **may not** discuss the exam with any other students until the exams have been returned.
- You may ask the instructor for clarification during the exam. Students who violate the Honor Code will be referred to the Honor Code Council.
- If you witness others violating the Honor Code, you have a duty to report them to the Honor Code Council.
- Students must pledge to obey the Honor Code by signing below. **Unsigned exams will not be graded.**

I understand and agree to abide by the principles of the Honor Code of Mount Holyoke College.

Signature *Date*

Multiple Choice (circle the letter corresponding to your answer) (4 pts each)

1. Suppose $Z \sim Normal(\mu = 0, \sigma = 1)$. Then, $P(Z = 1) =$
   (a) **0** - the probability of a single point for a continuous distribution is always 0.
   (b) 1
   (c) 0.5
   (d) 0.68

2. Complete the following sentence: When conducting a hypothesis test, we ____________ and then evaluate the test results to determine if there is enough evidence to ____________.
   (a) Assume that the null hypothesis is false; accept the null hypothesis
   (b) **Assume that the null hypothesis is true; reject the null hypothesis**
   (c) Assume that the alternative hypothesis is true; reject the null hypothesis
   (d) Assume the alternative hypothesis is false; reject the alternative hypothesis

3. Based on a random sample of 120 rhesus monkeys, a 95% confidence interval for the proportion of rhesus monkeys that live in a captive breeding facility and were assigned to research studies is (0.67, 0.83). Which of the following is true?
   (a) 95 of the sampled monkeys were assigned to research studies *(this is not the meaning of the 95% confidence interval)*
   (b) the margin of error for the confidence interval is 0.16 *0.83-0.16 is 0.67, so this can't be the margin of error*
   (c) if we used a different confidence level, the interval would not be symmetric about the sample proportion *(they are always symmetric)*
   (d) **none of the above are true**

4. The distribution of coin years (in circulation) is left-skewed - there are more newer coins in use than older coins. The sampling distribution for average coin year is

(d) **none of the above are true**

4. The distribution of coin years (in circulation) is left-skewed - there are more newer coins in use than older coins. The sampling distribution for average coin year is

   (a) left-skewed

   (b) right-skewed

   (c) **symmetric** (assuming the conditions of the Central Limit Theorem are satisfied.)

   (d) bimodal

5. When a variable follows a normal distribution, what percent of observations are contained within 1.96 standard deviations of the mean?

   (a) 90%

   (b) 68%

   (c) **95%** (1.96 is the critical value for a 95% confidence interval, and this problem describes how we get that value)

   (d) 99.7%

Short Answer

1. (20 pts) At yogurt factory, the amounts which go into yogurt containers are supposed to be normally distributed with mean 6 ounces and standard deviation 0.02 ounces (i.e. $X \sim \text{Normal}(6, 0.02)$). Once every 15 minutes, a container is selected from the production line and its contents are measured precisely. If the amount of yogurt is below 5.96 ounces or above 6.04 ounces, then the bottle fails quality control inspection.

   This is a Normal dist'n problem

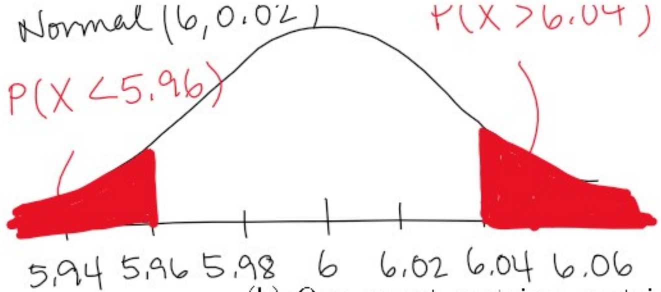
   For (a), you may need some or all of the following R output:

```
> pnorm(q=5.96, mean=6, sd=0.02)
[1] 0.02275013
> pnorm(q=5.96, mean=6, sd=0.02/sqrt(30))
[1] 3.163034e-28
```

   (a) What percent of yogurt containers fail quality control inspection? In addition to showing your notation and work, you may want to draw a picture. (5 pts)

Normal(6, 0.02) $P(X > 6.04)$ $P(\text{Failing}) = P(X>6.04) + P(X<5.96)$

$P(X<5.96)$ $= 2 \times P(X<5.96)$ (normal dist'n



$P(\text{failing}) = P(X > 6.04) + P(X < 5.96)$

$= 2 \times P(X < 5.96)$ (normal dist'n is symmetric)

$= 2 \times 0.02275$

$= 0.0455 \Rightarrow 4.55\%$ fail.

(b) One yogurt container contains 5.99 ounces. Calculate the z-score for this observation and interpret it. (6 pts)

$$Z = \frac{x - \mu}{\sigma} = \frac{5.99 - 6}{0.02} = -0.5$$

This yogurt container's contents are 0.5 standard deviations below the mean.

(c) Consider the sampling distribution for the average amount of yogurt in a container. Assume the sample size is 20.

Recall, if the observations come from a Normal(μ, σ), then the sampling dist'n for the mean is Normal$(\mu, \sigma/\sqrt{n})$.

(i) What is the mean of this sampling distribution? (3 pts)

6 ounces (same)

(ii) What is the standard error of this sampling distribution? (3 pts)

$\sigma/\sqrt{n} = 0.02/\sqrt{20}$

(iii) What is the shape of this sampling distribution? Is it approximate or exact? (3 pts)

Exactly normal (parent dist'n is normal)

2. (20 pts) A Gallup Poll found that 7% of teenagers (ages 13 to 17) suffer from arachnophobia and are extremely afraid of spiders. At a summer camp, there are 10 teenagers in each tent. Assume that these 10 teenagers are independent of each other.

(a) What distribution is most appropriate to model this problem? (2 pts)

Binomial

(b) What conditions need to be satisfied to apply the distribution you chose in (a)? Briefly identify how they are satisfied in this problem. (8 pts)

1. Fixed # trials, $n = 10$
2. Same probability of "success": 7% chance

1. Fixed # trials, $n = 10$
2. Same probability of "success": 7% chance suffer from arachnophobia
3. Each trial a success or failure (afraid or not)
4. Independent trials – assumed in problem.

For (c)-(d), you may need some or all of the following R output:

```
> dbinom(0, 10, 0.07)
[1] 0.4839823
> dbinom(1, 10, 0.07)
[1] 0.3642878
```

(c) Calculate the probability that at least one of them suffers from arachnophobia. (5 pts)

X = number who suffer from arachnophobia

$$P(X \geq 1) = 1 - P(X=0)$$
$$= 1 - 0.484$$

(d) Calculate the probability that at most one of them suffers from arachnophobia. (5 pts)

$$P(X \leq 1) = P(X=0) + P(X=1)$$
$$= 0.484 + 0.364$$
$$= 0.848$$

3. (20 pts) 400 students were randomly sampled from a large university, and 280 said they did not get enough sleep. In this problem, you will conduct a hypothesis test to check whether this represents a statistically significant difference from 50%. Use a significance level of 0.01.

(a) State whether the parameter of interest is a mean or a proportion. (1 pt)

(a) State whether the parameter of interest is a mean or a proportion. (1 pt)

proportion

(b) State the null and alternative hypotheses for your test. (4 pts)

$H_0: p = 0.5$

$H_A: p \neq 0.5$

(c) Check any conditions you need to satisfy to complete the test. (3 pts)

$np_0 = 400 \times 0.5 = 200 \geq 10$ ✓

$n(1-p_0) = 400 \times 0.5 = 200 \geq 10$ ✓

randomly sampled & independent

(d) Calculate the test statistic for your test. (4 pts)

$$z = \frac{\frac{280}{400} - 0.5}{\sqrt{0.5(1-0.5)/400}} = 8$$

(e) Estimate the p-value associated with your test statistic. (Hint: Use the 68-95-99.7 rule.) (1 pt)

The p-value is ≈ 0. Recall, since 99.7% of observations are within 3 standard deviations, only 0.3% is left in the tails. 8 is much farther away than 3.

(f) Interpret the result of your test and state your conclusion in the context of the problem. (7 pts)

There is very strong evidence against the null that 50% of students at this university do not get enough sleep. Rather, there is evidence that more than 50% do not get enough sleep (the z-score from (d) is positive).

4. (20 pts) An expensive restaurant claims that the average waiting time for dinner is approximately 60 minutes, but we **suspect that this claim is inflated** to make the restaurant appear more exclusive and successful. A random sample of 30 customers yields an average waiting time of 51 minutes. Assume the population standard deviation is 9.5 minutes.

   (a) State whether the parameter of interest is a mean or a proportion. (1 pt)

   mean

   For (b)-(c), you may need some or all of the following R output:

```
> qnorm(p=0.025, mean=0, sd=1, lower.tail=FALSE)
[1] 1.959964
> pnorm(q=0.025, mean=1, sd=1, lower.tail=FALSE)
[1] 0.8352199
> pnorm((51-60)/(9.5/sqrt(30)), 0, 1, lower.tail=TRUE)
[1] 1.057413e-07
```

   (b) Estimate the average waiting time for this restaurant. Be sure to check any relevant conditions and interpret your answer. (8 pts)

   Construct a confidence interval

   $\bar{x} \pm z^* \times SE = 51 \pm 1.96 \times \frac{9.5}{\sqrt{30}}$

   $= (47.6, 54.4)$ minutes

   We are 95% confident that the true mean wait time is between 47.6 and 54.4 minutes.

   (c) Conduct a hypothesis test to determine if there is evidence that the reported wait time is inflated (i.e. the true wait time is less than reported). State your conclusion in the context of the problem. (8 pts)

   ① State hypotheses: $H_0: \mu = 60$ vs. $H_A: \mu < 60$

   ② Use confidence interval or z score + p-value

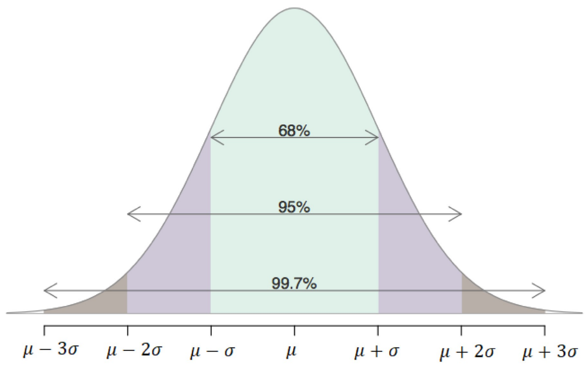
   ③ Interpret: Since our 95% CI is below 60 minutes, there is evidence that the mean wait time is less than 60 minutes

   (d) If you were to construct a 90% confidence interval that corresponded to this hypothesis test would you expect 60 minutes to be in the interval? (3 pts)

   No, since it was not in the 95% confidence interval, it would not be in the 90% confidence interval because it is narrower.

STAT 140 Midterm II Formula Sheet

- 68-95-99.7 Rule



- Z score: $z = \frac{x-\mu}{\sigma}$
- Binomial mean: $\mathrm{E}(X) = np$
- Binomial standard deviation: $\sqrt{\mathrm{Var}(X)} = \sqrt{np(1-p)}$
- Binomial formula: $P(X=k) = \underbrace{\binom{n}{k}}_{\#\ scenarios} \underbrace{p^k(1-p)^{n-k}}_{P(single\ scenario)} = \frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$
- Standard errors

  $SE = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$

  $SE = \sqrt{\frac{p(1-p)}{n}} \approx \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$
- Conditions for "sufficiently large" sample size

  $np \geq 10$ and $np(1-p) \geq 10$

  $n \geq 30$
- Confidence intervals

  point estimate ± margin of error = point estimate ± critical value × SE

  $\hat{p} \pm z^* \times SE$

  $\bar{x} \pm z^* \times SE$

  $\bar{x} \pm t^*_{df} \times SE$; $df = n - 1$
- Hypothesis tests

  test statistic $= \frac{\text{point estimate} - \text{hypothesized value}}{\text{standard error}}$

  $z = \frac{\widehat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$

  $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$

  $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$

$z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$

$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$

- R Code: Normal Distribution

  $P(X < x) = \texttt{pnorm}(\texttt{q} = x, \texttt{mean} = \mu, \texttt{sd} = \sigma, \texttt{lower.tail} = \texttt{TRUE})$

  $P(X > x) = \texttt{pnorm}(\texttt{q} = x, \texttt{mean} = \mu, \texttt{sd} = \sigma, \texttt{lower.tail} = \texttt{FALSE})$

  $x = \texttt{qnorm}(\texttt{p} = P(X < x), \texttt{mean} = \mu, \texttt{sd} = \sigma, \texttt{lower.tail} = \texttt{TRUE})$

  $x = \texttt{qnorm}(\texttt{p} = P(X > x), \texttt{mean} = \mu, \texttt{sd} = \sigma, \texttt{lower.tail} = \texttt{FALSE})$

- R Code: t-Distribution

  $P(T < t) = \texttt{pt}(\texttt{q} = t, \texttt{df} = n - 1, \texttt{lower.tail} = \texttt{TRUE})$

  $P(T > t) = \texttt{pt}(\texttt{q} = t, \texttt{df} = n - 1, \texttt{lower.tail} = \texttt{FALSE})$

  $t = \texttt{qt}(\texttt{p} = P(T < t), \texttt{df} = n - 1, \texttt{lower.tail} = \texttt{TRUE})$

  $t = \texttt{qt}(\texttt{p} = P(T > t), \texttt{df} = n - 1, \texttt{lower.tail} = \texttt{FALSE})$

- R Code: Binomial Distribution

  $P(X = k) = \texttt{dbinom}(\texttt{x} = k,\ \texttt{size} = n,\ \texttt{prob} = p)$

  $P(X \leq k) = \texttt{pbinom}(\texttt{q} = k,\ \texttt{size} = n,\ \texttt{prob} = p,\ \texttt{lower.tail} = \texttt{TRUE})$

  $P(X > k) = \texttt{pbinom}(\texttt{q} = k,\ \texttt{size} = n,\ \texttt{prob} = p,\ \texttt{lower.tail} = \texttt{FALSE})$