Review

Exam 2 Review

Wendy Wang
wwang60@central.uh.edu

Exam 2 - 3339

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Outline

Chapter 5 Continuous Random Variables

dirivative

- Density functions

Pdf $\to f(x)$
Cdf $\to F(x)$

- Know how to calculate expected values and quantiles for continuous distributions.

$E(x) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx$

- Named distributions
  - Uniform
  - Exponential
  - Gamma
  - Normal

$\boxed{\text{none}}$

$Var(x) = E(x^2) - (E(x))^2$

$= \int_{-\infty}^{\infty} x^2 f(x)\,dx - [E(x)]^2$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Chapter 6 Sampling Distributions

- Expected values and variances of X + Y
- Applying the Central Limit Theorem
- The sampling distribution of the sample means, $\bar{X}$.
- The sampling distribution of the proportions, $\hat{p}$.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Confidence Intervals Chapter 7

- point estimation $\bar{x}$, $\hat{p}$
- CI

Know how to calculate the confidence intervals for mean (μ), proportions (p) and standard deviation (σ).

- For mean if σ, population standard deviation is given: $\bar{x} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$.
- For mean if σ is not given: $\bar{x} \pm t_{\alpha/2,df}\left(\frac{s}{\sqrt{n}}\right)$.
- For proportions $\hat{p} = \frac{X}{n}$: $\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.
- For standard deviation: $lcl = \sqrt{\frac{(n-1)s^2}{qchisq(1-\alpha/2,n-1)}}$ and $ucl = \sqrt{\frac{(n-1)s^2}{qchisq(\alpha/2,n-1)}}$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Confidence Intervals

width $= 2 \times ME$
$= 2 \times \left(\frac{z_{\alpha/2}}{\sqrt{n}} \cdot \sigma \right)$

as $\sigma \uparrow$ width $\uparrow$
as $C \uparrow$, width $\uparrow$
as $n \uparrow$ width $\downarrow$

- Know how the confidence interval changes as the confidence level C changes and as the sample size changes.
- Know how to interpret confidence intervals.
- Know how to determine a sample size given confidence level and margin of error.
  - For means $\boxed{n \geq \left(\frac{z_{\alpha/2} \times \sigma}{E}\right)^2}$, where E = margin of error.
  - For proportions $n = p^*(1 - p^*)(\frac{z_{\alpha/2}}{E})^2$, where p^* is some previous knowledge of the proportion if not known we use 0.5.

Hypothesis Tests Chapter 8

H_0:

H_a:

- Free response

- Know how to set up null and alternative hypotheses.
- Be able to determine a rejection region, given the level of significance (α).
- Calculate a test statistic.

	$H_0 : \mu = \mu_0$ if σ is given	$H_0 : \mu = \mu_0$ if σ is unknown	$H_0 : p = p_0$
Test Statistic	$z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$	$t = \frac{\bar{x}-\mu_0}{s\sqrt{n}}$	$z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

- Be able to calculate the p-value and know to reject (RH0) or fail to reject (FRH0).
- Know the difference between Type I and Type II errors.
- Be able to make a conclusion in context of the problem.

Type I error only occur when we reject H_0.

UNIVERSITY of HOUSTON DEPARTMENT OF MATHEMATICS

- Matched paired t-test:
  - when our data samples are DEPENDENT upon one another (like before and after results).
  - Hypotheses - $H_0 : \mu_d = 0$ and $H_a : \mu_d \neq 0$ or $\mu_d < 0$ or $\mu_d > 0$. Where μ_d is the mean of the differences.
- Comparing two means
  - Each group is considered to be a sample from a distinct population.The responses in each group are independent of those in the other group.
  - Hypotheses - $H_0 : \mu_1 = \mu_2$ and $H_a : \mu_1 \neq \mu_2$ or $\mu_1 < \mu_2$ or $\mu_1 > \mu_2$.

Two sample proportions test will NOT be tested by test 2.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

- When we use the sample standard deviations we use the **two-sample t statistic**

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

with degrees of freedom approximated by the smaller value of $n_1 - 1$ or $n_2 - 1$, or based on the formula provided (check test 2 formula card).

min(n_1-1, n_2-1)

qt$\left(\frac{1+C}{2}, df\right)$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

What to Expect on the Exam

The test has two parts

1. multiple choice questions.
2. free responses questions.

$7 \times 7 = 49$

$3 \times 17 = 51$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example 1 (from quiz 11)

The average height of students at UH from a SRS of 17 students gave a standard deviation of 2.9 feet. Construct a 95% confidence interval for the standard deviation of the height of students at UH. Assume normality for the data. var formula then sqrt

$n = 17$

$s^2 = 2.9^2$

$$\sqrt{\frac{(17-1)\times 2.9^2}{qchisq(\frac{1+c}{2}, 17-1)}} , \sqrt{\frac{(17-1)\times 2.9^2}{qchisq(\frac{\alpha}{2}, 17-1)}}$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example 2

Identify the most appropriate test to use for the following situation:

- A national computer retailer believes that the average sales are greater for salespersons with a college degree. A random sample of 14 salespersons with a degree had an average weekly sale of \$3542 last year, while 17 salespersons without a college degree averaged \$3301 in weekly sales. The standard deviations were \$468 and \$642 respectively. Is there evidence to support the retailer's belief?

  n_1 n_2 $\bar{x}_1 =$ $\bar{x}_2 =$

  $s_1 = 468$ $s_2 = 642$ Two sample t test

- In low-speed crash test of five BMW cars, the repair costs were computed for a factory-authorized repair center and an independent repair facility. The results are as follows.

Authorized repair center	\$797	\$571	\$904	\$1147	\$418
Independent repair center	\$523	\$488	\$875	\$911	\$297

We want to estimate the mean of the difference between the two repair centers.

Example 3 Formula Not provided

A population is normally distributed with unknown mean and standard deviation 2.9. A sample is collected and a 93% confidence interval for the population mean is constructed. If the confidence interval is given below, what is the sample size? (43.503258,44.716742)

CL ME

$$n > \left(\frac{z_{\frac{\alpha}{2}} \cdot \sigma}{ME}\right)^2$$

$$= \left(\frac{qnorm\left(\frac{1.93}{2}\right) \times 2.9}{}\right)$$

Example 4

Suppose a sample of 100 subjects was taken and their scores on an exam recorded. If the population mean for the exam is 67 and population variance is 36, $\sigma^2 = 36 \to \sigma = 6$

1. What is the mean and standard error of the sampling distribution, $\bar{X}$? $\mu_{\bar{x}} = \mu$ $\quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ $\quad \frac{6}{\sqrt{100}} = 0.6$
2. find $P(\boxed{\bar{X}} < 70)$.
3. find $P(45 < \bar{X} < 74)$.

2\. CLT $n > 30 \to \bar{x}$ $\underline{\text{approx}}$ $N(\mu_{\bar{x}} = 67, \sigma_{\bar{x}} = 0.6)$

$P(\bar{x} < 70) = \text{pnorm}(70, 67, 0.6)$

3\. $P(45 < \bar{x} < 74) = P(\bar{x} < 74) - P(\bar{x} < 45)$

$= \text{pnorm}(74, 67, 0.6) - \text{pnorm}(45, 67, 0.6)$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example 5

A simple random sample of 100 8th graders at a large suburban middle school indicated that 86% of them are involved with some type of after school activity. Find the 98% confidence interval that estimates the proportion of them that are involved in an after school activity.

a) (0.679, 0.891)
b) (0.699, 0.941)
c) (0.779, 0.941)
d) (0.829, 0.834)
e) (0.779, 0.741)

proportion formula

$$\hat{p} \pm z * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$n = 100$

$\hat{p} = 0.86$

always $\left(\frac{1+C}{2}\right)$

$$\text{qnorm}\left(\frac{1+C}{2}\right)$$

$$= \text{qnorm}\left(\frac{1.98}{2}\right)$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example 6

Suppose a continuous random variable X has the following pdf:

$$f(x) = \begin{cases} \frac{1}{a}x & 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

$\int_{-\infty}^{\infty} f(x)dx = 1$

$\to a =$

1. Find the value of a that makes this function a valid pdf.
2. Determine the expected value of X. $E(X) = \int_{-\infty}^{\infty} x f(x) dx =$
3. Find $P(X \le 1)$ $\int_{-\infty}^{1} f(x) dx$
4. Determine the CDF.
5. Find the value of c such that $P(X \le c) = 0.04$.
6. Determine Q_3, the third quartile.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example 7

population μ

Bags of a certain brand of tortilla chips claim to have a net weight of 14 ounces. Net weights actually vary slightly from bag to bag and are normally distributed with mean μ. A representative of a consumer advocate group wishes to see if there is any evidence that the mean net weight is less than advertised. Test this claim at the 5% significance level. To do this test, he selects 16 bags of this brand at random and determines the net weight of each. From this sample, he finds the mean to be $\bar{x} = 13.8668$ and the standard deviation to be $s = 0.25$.

$H_0: \mu = 14$ $\quad$ $H_a: \mu < 14$

1. Determine the null and alternative hypothesis.
2. Calculate the test statistic for this significance test.
3. Give the rejection region.
4. Determine the p-value
5. State the conclusion in context of the problem.

no σ use t

$$t = \frac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n}}\right)} = \frac{13.86}{\frac{0.35}{\sqrt{16}}} = 2.13$$

$\alpha = 0.05$

cv $\quad$ qt(0.05, 16-1) $= -1.75$

-2.13 is in Rejection Region. Reject H_0.

$$P\text{-value} = P(t \le -2.13)$$
$$= pt(-2.13, 16-1) = 0.025 < \alpha$$

will reject

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example 8

The random variable X has cumulative distribution function:

$$F(x) = \begin{cases} 0 & x < -4 \\ \frac{1}{8}x + \frac{1}{2} & -4 \le x \le 4 \\ 1 & x > 4 \end{cases}$$

What is the density function? What is the Q1, Q2, and Q3 of this random variable?

$$f(x) = F'(x) = \begin{cases} 0 & x < -4, \text{ or } x > 4 \\ \frac{1}{8} & -4 \le x \le 4 \end{cases}$$

$E(x) = ?$ $Var(x) = ?$

$P(x \le Q_3) = 0.5$

Q_1 $P(x \le Q_1) = 0.25 \to F(Q_1)$ for 0.25 solve for Q_1

$P(x \le Q_2) = 0.50$

Example 9

On review

A sample of 97 Duracell batteries produces a mean lifetime of 10.40 hours and standard deviation 4.83 hours. A sample of 148 Energizer batteries produces a mean lifetime of 9.26 hours and a standard deviation of 4.68 hours. At a 5% significance level, can we assert that the average lifetime of Duracell batteries is greater than the average lifetime of Energizer batteries?

1. State an appropriate null hypothesis and alternative hypothesis.
2. Carry out the test. Give the p-value, and then interpret the result.
3. Give a 90% confidence interval for the difference.

$H_0: \mu_1 = \mu_2$ $\quad$ $H_a: \mu_1 > \mu_2$

$$(\bar{x}_1 - \bar{x}_2) \pm t \times \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example 10

matched paired t test

We want to determine a 95% confidence interval for the difference in the repair cost of the authorized repair center and the independent repair center.

Authorized repair center	\$797	\$571	\$904	\$1147	\$418
Independent Repair center	\$523	\$488	\$875	\$911	\$297
Differences	\$274	\$83	\$29	\$236	\$121

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Inference for Matched Pairs

- The previous question is a matched pair.
- We are looking at the same car. The subject units are exactly the same for both responses.
- We calculate the differences first and find the mean and standard deviation of the differences.
- Then this problem is the same as a one-sample confidence interval.
  - We first find the differences from each observation.
  - The point estimate is $\bar{x}_d$ = mean of the differences.
  - The standard deviation is s_d = the standard deviation of the differences.
  - Then the margin of error is $m = t^* \left(\frac{s_d}{\sqrt{n}}\right)$.
  - The confidence interval is $\boxed{\bar{x}_d \pm t^* \left(\frac{s_d}{\sqrt{n}}\right).}$
- If we want a hypothesis test, the test statistic is: $t = \frac{\bar{x}_d - \mu_d}{s_d / sqrtn}$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

R code

```
> auth=c(797,571,904,1147,418)
> indep=c(523,488,875,911,297)
> t.test(auth,indep,conf.level = 0.95, paired = TRUE)
```

(conf. level=.99, paired=TRUE)

```
Paired t-test

data:  auth and indep
t = 3.2148, df = 4, p-value = 0.03244
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
20.26155 276.93845
sample estimates:
mean of the differences
148.6
```

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Full steps to perform a hypothesis test

two sample test

☆ on free response Q6, Q8

Full steps to perform a one-mean hypothesis test:

1. State your null hypothesis and alternative hypothesis
2. Sketch the reject region
3. Calculate the test statistic and denote the statistic on the plot above
4. Determine the P-value for your test
5. Make your conclusion

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

What You Need and What is Provided

- Provided
  - Rstudio; it will be a link you see in the exam.
  - Formula sheet; it will be a link you see in the exam. You can not bring a printed copy to the exam.
  - CASA online calculator; it will be a link you see in the exam. You can not use your own calculator during the exam.
  - Scratch paper and answering booklet for written questions will be provided by CASA testing center.
- Can bring
  - Pencil
  - Your Cougar Card or other photo ID (see CASA instruction)

Questions?