Exam 1 Review

Review

Wendy Wang
wwang60@central.uh.edu

Test 1 Review

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Dates of Exam and Covered Chapters

- Dates: March 2-March 4

poissa lm(y~x)

- Chapters: 1, 2, 3 and 4 (4.1-4.8) and section 9.1-9.2
- ~~Jointly distributed variables~~ will not be covered by Test 1

4.9

Chapter 1

- Sample - Simple random sample
- Population versus sample
- Parameter versus statistic.
- Categorical variables
- Quantitative variables; continuous and discrete

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

- Describing distributions with numbers and graphs
- Center - mean, median, mode
- Spread - range, interquartile range (IQR), standard deviation
- Location - percentiles, quartiles (Q_1 and Q_3), $1.5 \times$ IQR

outliers

- The five number summary

min Q_1 Q_2 Q_3 max

25% 50% 75%

median of first half

median of set

median of 2nd half

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

- Describing distributions with graphs
- Categorical variables - bar chart, pie chart
- Quantitative variables - histogram, stemplot, boxplot

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Notation	Description
$a \in A$	The object *a* is an element of the set *A*.
$A \subseteq B$	Set *A* is a subset of set *B*. That is every element in *A* is also in *B*.
$A \subset B$	Set *A* is a proper subset of set *B*. That is every element that is is in *A* is also in set *B* and there is at least one element in set *B* that is no in set *A*.
$A \cup B$	A set of all elements that are in *A* **or** *B*.
$A \cap B$	A set of all elements that are in *A* **and** *B*.
U	Called the **universal set**, all elements we are interested in.
A^C	The set of all elements that are in the universal set but are not in set A.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Chapter 3 section 2: Counting Techniques

- If an experiment can be described as a sequence of k steps with n_1 possible outcomes on the first step, n_2 possible outcomes on the second step, and so on, then the total number of experimental outcomes is given by $(n_1)(n_2)\dots(n_k)$.
- **Permutations**: allows one to compute the number of outcomes when r objects are to be selected from a set of n objects where the order of selection is important. The number of permutations is given by $P_r^n = \frac{n!}{(n-r)!}$
- When we allow repeated values, The number of orderings of n objects taken r at a time, with repetition is n^r.
- The number of permutations, P, of n objects taken n at a time with r objects alike, s of another kind alike, and t of another kind alike is $P = \frac{n!}{r!s!t!}$
- The number of circular permutations of n objects is $(n-1)!$.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Combiniations

Combinations counts the number of experimental outcomes when the experiment involves selecting r objects from a (usually larger) set of n objects. The number of combinations of n objects taken r unordered at a time is

$$C_r^n = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

For any event A, the probability of A is

$$P(A) = \frac{\text{number of times A occurs}}{\text{total number of outcomes}}.$$

General Rules of Probability $1 \geq P \geq 0$

1. The probability $P(A)$ of any event A satisfies $0 \leq P(A) \leq 1$.
2. If S is the sample space in a probability model, then $P(S) = 1$.
3. **Complement rule**: For any event A,

$$P(A^C) = 1 - P(A)$$

4. **General rule for addition**: For any two events A and B

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

5. **General rule for multiplication**: For any two events A and B

$$P(A \cap B) = P(A) \times P(B, \text{ given } A)$$

$P(A) \times P(B|A)$

or

$$P(A \cap B) = P(B) \times P(A, \text{ given } B)$$

$= P(B) \times P(A|B)$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Chapter 3 section 5: General Probability Models

- Two events are disjoint if the occurrence of one prevents the other from happening.
- If two events A and B are disjoint then $P(A \text{ and } B) = P(A \cap B) = 0$.
- Two events are independent if the occurrence of one does not change the *probability* of the other.
- If two events A and B are independent then

$P(A|B) = P(A)$

$$P(A \text{ and } B) = P(A \cap B) = P(A) \times P(B).$$

- **Conditional Probability**: For any two events A and B, the probability of A given B is

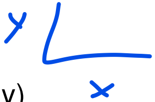
$$P(A \text{ given } B) = P(A|B) = \frac{P(A \cap B)}{P(B)}$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

- Define a discrete random variable.
- Binomial distribution; Hypergeometric distriution; Possion distribution.
- Know how to calculate the probability for different distributions.
- Know how to calculate the expected values and variances from a discrete probability distribution.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Chapter 9: Least-Squares Regression



- Scatterplot of two variables, in R plot(x,y)
- Finding covariance, in R cov(x,y), correlation, in R cor(x,y) and LSLR, in R lm(y x).
  least square linear regression lm(y~x)
- Understand what each of these numbers mean.

What is on the Exam

- 75 minutes
- 7 multiple choice questions (7 points each)
- 3 free response questions (17 points each)
- Answer the 7 multiple choice problems by selecting your answer on the computer at CASA testing center
- Write your answers to the free response problems on the answering booklet (The booklet will be provided by CASA testing center).
- You will be provided with links of Test 1 formula, casa online calculator, and Rstudio.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Possible Multiple Choice Questions

- Detecting outliers.
- Looking at the graphs and determine the shape.
- Know the difference between the types of variables.
- Using the probability rules.
- Know how to find probabilities from a discrete distribution table and binomial/Poisson/hypergeometric distribution
- Know how to find expected value and variance from a discrete distribution table and binomial/Poisson/hypergeometric distribution and from a linear expression.

Possible Free Response Questions

- Determining outliers, shape, center and spread from descriptive numerical values.
- Probability rules, including the Baye's rule.
- Know when two events are independent.
- Know how to find probabilities from a discrete distribution table.
- Know how to find expected value from a discrete distribution table.
- Creating a least-squares linear equation from the data, scatterplot, correlation, coefficient of determination, and residual.

Example 1

$X \sim \text{hype}(m, n, k)$ n,p

An urn has 20 blue marbles and 15 red marbles in it. Determine the probability that if 5 marbles are selected, at least two will be blue.

X = # of blue marble from 5

$X \sim \text{hypegeom}(m=20, n=15, k=5)$

m = 20 blue
n = 15 red

35

k = 5 sample

calculate each & add up

P(X at least 2) = P(X ≥ 2) = P(2, 3, 4, 5)

or

= 1 − phyper(1, 20, 15, 5)

largest possible value less than or equal to.

P(X at most 4)
= P(X ≤ 4)
phyper(4, 20, 15, 5)

P(X fewer than 2)

Example 2

Suppose you have a distribution, X, with mean = 6 and standard deviation = 8. Define a new random variable $Y = 9X - 3$. Find the mean and standard deviation of Y.

$\mu_x = 6 \quad \sigma_x = 8$

$E(Y) = E(9x-3)$

$= 9E(x) - 3 = 9 \times 6 - 3 =$

$\sigma_Y = 9\sigma_x = 9 \times 8 = 72$

$Var(Y) = \sigma_Y^2 = (9\sigma_x)^2 = 81 \times 8^2$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example 3

Newsweek in 1989 reported that 60% of young children have blood lead levels that could impair their neurological development. Assuming a random sample from the population of all school children at risk, find the probability that at least 5 children out of 10 in a sample taken from a school may have a blood level that may impair development.

x = # of children who may have a blood level that may impair development

$x \sim \text{binom}(n=10, p=0.6)$

$P(x \text{ at least } 5) = P(x \geq 5)$

$= 1 - P(x < 5) = 1 - \text{pbinom}(4, 10, 0.6)$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example 4: class example

Suppose the random variable X takes on possible values x = 0, 1, 2, 3 and has pmf given by f(x) = (x+1)/k , determine the value of k.

probability mass function

or $P(X<2)$

x	0	1	2	3
pmf	$\frac{0+1}{k} = \frac{1}{k}$	$\frac{1+1}{k} = \frac{2}{k}$	$\frac{2+1}{k} = \frac{3}{k}$	$\frac{3+1}{k} = \frac{4}{k}$

$$\frac{1}{k} + \frac{2}{k} + \frac{3}{k} + \frac{4}{k} = 1$$

solve for k

$$\frac{10}{k} = 1$$

$$10 = k$$

Example 5

$X \sim \text{Poisson}(\text{mean} = \mu = 5/\text{hr})$

The number of people arriving for treatment at an emergency room can be modeled by a Poisson process with a mean of five people per hour.

1. What is the probability that exactly four arrivals occur at a particular hour?

   $P(X=4)$ $\quad x \geq 0$

   $= \text{dpois}(4, 5)$

   $\boxed{\text{4. P(at least 4 ppl arrive during a 2 hrs.) on phone}}$

2. What is the probability that at least four people arrive during a particular hour?

   $P(X \text{ at least } 4) = P(X \geq 4) = 1 - P(X < 4)$

   $= 1 - \text{ppois}(3, 5)$

3. How many people do you expect to arrive during a 45-min period?

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example 6

Suppose that for events A and B, $P(A) = 0.4$, $P(B) = 0.3$, $P(A \bigcup B) = 0.5$

a. Compute $P(A|B)$

b. Are events A and B independent?

a. $P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{0.2}{0.3}$

$\underset{0.5}{P(A \cup B)} = \underset{.4}{P(A)} + \underset{.3}{P(B)} - \underset{?\ .2}{P(A \cap B)}$

b. $P(A|B) \neq P(A) \rightarrow$ A and B are NOT

$P(A \cap B) \neq P(A) \times P(B) \rightarrow$ A and B are not

Example 7

Answer the questions below with the following data set:

x	2	8	8	13	16	19
y	22	29	28	40	33	41

x=c( )
y=c( ) plot(x,y)

1. Create a scaterplot from the data.
2. What is the correlation coefficient? cor(x,y)
3. Determine the coefficient of determination. (cor(x,y))^2
4. Develop a LSLR for the given data. lm(y~x)
5. Give the residual value for x = 13. residual= observed y − predicted y

= 40 − (20.479 + 0.1067 * 13)

6. is the fitted line good or bad?

$r^2 = 0.8003$ good is close to 1.

$0 < R^2 < 1$ 0.6 is bad

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example 8: Bayesian question in hw2

Example 9: class example

Example: Let X have pmf. given by

x	1	2	3	4
f(x)	0.4	0.2	0.3	0.1

Determine $E[X]$, $E[X^2]$, $Var[X]$ and the standard deviation of X.

$E(X) = \sum x \cdot P(X=x)$

$= 1 \times 0.4 + 2 \times 0.2 + 3 \times 0.3 + 4 \times 0.1$

$E(X^2) = \sum x^2 P(X=x)$

$= 1^2(0.4) + 2^2(0.2) + 3^2(0.3) + 4^2(0.1)$

$var(X) = E(X^2) - E(X)^2$

Example 10: hw3

4.

X	-1	0	1	2	3
$P(X)$	m	n	1.2	0.2	0.2

a. $E(X) = \sum x P(X=x) = -1 \times m + 0 \times n + 1 \times 0.2$
$+ 2 \times 0.1 + 3 \times 0.2 = 0.6$

$m = ?$ $m + n + 0.2 + 0.1 + 0.2 = 1$

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)

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Questions?