Chi-Square Tests & Final Exam Review

Chapter 12

Cathy Poliak, Ph.D.
cathy@math.uh.edu
Office Fleming 11c

Department of Mathematics
University of Houston

Lecture 14 - 3339

Outline

1. Goodness of Fit Tests
2. Chi-Square
3. Examples
4. χ^2 Test of Independence
5. Final Exam Review
6. Examples

Candy

Mars Inc. claims that they produce M&Ms with the following distributions:

Brown	30%	Red	20%	Yellow	20%
Orange	10%	Green	10%	Blue	10%

A bag of M&Ms was randomly selected from the grocery store shelf, and the color counts were:

Brown	14	Red	14	Yellow	5
Orange	7	Green	6	Blue	10

We want to know if the distribution of color the same as the manufacturer's claim.

Goodness-of-fit Test

- This is a test to see how well on sample proportions of categories "match-up" with the known population proportions.
- The Chi-square goodness-of-fit test extends inference on proportions to more than two proportions by enabling us to determine if a particular population distribution has changed from a specified form.
- Hypotheses:
  - H_0: The proportions are the same as what is claimed.
  - H_a: At least one proportion is different as what is claimed.

  This would be better in context of the problem. For example in our M&Ms test;
  - H_0: The distribution of candy colors is as the manufacturer claims.
  - H_a: The distribution of candy colors is not what the manufacturer claims.

Goodness-of-fit Test

- This is a test to see how well on sample proportions of categories "match-up" with the known population proportions.
- The Chi-square goodness-of-fit test extends inference on proportions to more than two proportions by enabling us to determine if a particular population distribution has changed from a specified form.
- Hypotheses:
  - H_0: The proportions are the same as what is claimed.
  - H_a: At least one proportion is different as what is claimed.

  This would be better in context of the problem. For example in our M&Ms test;
  - H_0: The distribution of candy colors is as the manufacturer claims.
  - H_a: The distribution of candy colors is not what the manufacturer claims.

Goodness-of-fit Test

- This is a test to see how well on sample proportions of categories "match-up" with the known population proportions.
- The Chi-square goodness-of-fit test extends inference on proportions to more than two proportions by enabling us to determine if a particular population distribution has changed from a specified form.
- Hypotheses:
  - H_0: The proportions are the same as what is claimed.
  - H_a: At least one proportion is different as what is claimed.

  This would be better in context of the problem. For example in our M&Ms test;
  - H_0: The distribution of candy colors is as the manufacturer claims.
  - H_a: The distribution of candy colors is not what the manufacturer claims.

Goodness-of-fit Test

- This is a test to see how well on sample proportions of categories "match-up" with the known population proportions.
- The Chi-square goodness-of-fit test extends inference on proportions to more than two proportions by enabling us to determine if a particular population distribution has changed from a specified form.
- Hypotheses:
  - H_0: The proportions are the same as what is claimed.
  - H_a: At least one proportion is different as what is claimed.

  This would be better in context of the problem. For example in our M&Ms test;
  - H_0: The distribution of candy colors is as the manufacturer claims.
  - H_a: The distribution of candy colors is not what the manufacturer claims.

Chi-Square Test

Test Statistic: Called the **chi-square statistic** is a measure of how much the observed cell counts diverge from the expected cell counts. To calculate for each problem you will make a table with the following headings:

Observed Counts (O)	Expected Counts (E)	$\frac{(O-E)^2}{E}$

The sum of the third column is called the Chi-square test statistic, χ^2.

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

Where expected counts = total count $\times$ proportion of each category.

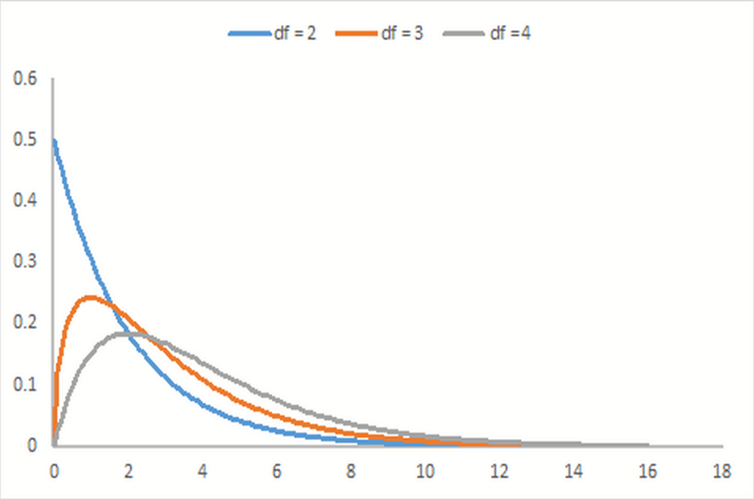
Chi-square of M&Ms

Color	Observed Counts (O)	Proportions	Expected Counts (E)	$\frac{(O-E)^2}{E}$
Brown	14	0.3		
Red	14	0.2		
Yellow	5	0.2		
Orange	7	0.1		
Green	6	0.1		
Blue	10	0.1		

Chi-square

- Chi-square distributions have only positive values and are skewed right.
- This has a degrees of freedom which is $n - 1$.
- As the degrees of freedom increases it become more like a Normal distribution.
- The total area under the χ^2 curve is 1.
- To find area under the curve
  - Table provided
  - In R: 1 - pchisq(x,df)

Chi-Square



Assumptions for a Chi-Square Goodness-of-fit Test

1. The sample must be an SRS from the populations of interest.
2. The population size is at least 10 times the size of the sample.
3. All expected cell counts must be at least 5.

Is the manufacturers claim correct?

Using R

- chisq.test(c(list of observed values),correct = FALSE, p = c(list of proportions))
- If we are not given a list of proportions then p = 1/n and that is a default for R so we do not need to give that information.

```
> chisq.test(c(14,14,5,7,6,10),correct=FALSE,p=c(.3,.2,.2,
.1,.1,.1))

Chi-squared test for given probabilities

data:  c(14, 14, 5, 7, 6, 10)
X-squared = 8.4345, df = 5, p-value = 0.1339
```

Using R

- chisq.test(c(list of observed values),correct = FALSE, p = c(list of proportions))
- If we are not given a list of proportions then p = 1/n and that is a default for R so we do not need to give that information.

```
> chisq.test(c(14,14,5,7,6,10),correct=FALSE,p=c(.3,.2,.2,
.1,.1,.1))

Chi-squared test for given probabilities

data:  c(14, 14, 5, 7, 6, 10)
X-squared = 8.4345, df = 5, p-value = 0.1339
```

Zodiac Signs

Does your zodiac sign determine how successful you will be in later life? *Fortune* magazine collected the zodiac signs of 256 heads of the largest 400 companies. The following are the number of births for each sign:

Sign	Births
Aries	23
Taurus	20
Gemini	18
Cancer	23
Leo	20
Virgo	19
Libra	18
Scorpio	21
Sagittarius	19
Capricorn	22
Aquarius	24
Pisces	29

From: *Intro Stats*, De Veaux, Velleman, Bock. 2nd Edition, Pearson, pg 604.

Example

The following table shows three different airlines **row variable** and the number of delayed or on-time flights **column variable** from lightstats.com.

	Delayed	On-time	Total
American	112	843	955
Southwest	114	1416	1530
United	61	896	957
Total	287	3155	3442

- Does on-time performance depend on airline?
- We will use a significance test to answer this question.

Significance Tests For Two-Way Tables

1. The assumptions necessary for the test to be valid are:
   a. The observations constitutes a simple random sample from the population of interest, and
   b. The expected counts are at least 5 for each cell of the table.

2. Hypotheses
   - Null hypothesis: There is no association (independence) between the row variable and column variable.
   - Alternative hypothesis: There is an association (dependence) between the row variable and column variable.
   - In the previous example:

$$H_0 : \text{Airline and on-time performance are independent.}$$
$$H_A : \text{On-time performance depends on airline.}$$

Significance Tests For Two-Way Tables

3. Test Statistic: Called the **chi-square statistic** is a measure of how much the observed cell counts in a two-way table diverge from the expected cell counts. To calculate.

$$X^2 = \sum \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}}$$

Where "observed" represents an observed sample count, and "expected" is calculated by

$$\text{expected count} = \frac{\text{row total} \times \text{column total}}{n}$$

The sum is over all $r \times c$ cells in the table. Where r is the number of rows and c is the number of columns.

Significance Tests For Two-Way Tables

If H_0 is true, the chi-square statistic X^2 has approximately a χ^2 distribution with $(r-1)(c-1)$ degrees of freedom. Where r = number of rows and c = number of columns.

4. The P-value for the chi-square test is $P(\chi^2 \geq X^2)$. Given that all of the expected cell counts be 5 or more.

5. Decision: If P-value is less than α level of significance, we reject H_0. Otherwise we fail to reject H_0.

6. Conclusion: In context of the problem.

Example

The following table shows three different airlines **row variable** and the number of delayed or on-time flights **column variable** from flightstats.com.

	Delayed	On-time	Total
American	112	843	955
Southwest	114	1416	1530
United	61	896	957
Total	287	3155	3442

Does on-time performance depend on airline?

Expected cell counts

The following table gives the expected cell count.

	Delayed	On-time	Total
American	$\frac{955\times287}{3442}=79.6296$	$\frac{955\times3155}{3442}=875.3704$	955
Southwest	$\frac{1530\times287}{3442}=127.5741$	$\frac{1530\times3155}{3442}=1402.4259$	1530
United	$\frac{957\times287}{3442}=79.7963$	$\frac{957\times3155}{3442}=877.20367$	957
Total	287	3155	3442

Expected cell counts

The following table gives the expected cell count.

	Delayed	On-time	Total
American	$\frac{955\times287}{3442} = 79.6296$	$\frac{955\times3155}{3442} = 875.3704$	955
Southwest	$\frac{1530\times287}{3442} = 127.5741$	$\frac{1530\times3155}{3442} = 1402.4259$	1530
United	$\frac{957\times287}{3442} = 79.7963$	$\frac{957\times3155}{3442} = 877.20367$	957
Total	287	3155	3442

Expected cell counts

The following table gives the expected cell count.

	Delayed	On-time	Total
American	$\frac{955\times287}{3442} = 79.6296$	$\frac{955\times3155}{3442} = 875.3704$	955
Southwest	$\frac{1530\times287}{3442} = 127.5741$	$\frac{1530\times3155}{3442} = 1402.4259$	1530
United	$\frac{957\times287}{3442} = 79.7963$	$\frac{957\times3155}{3442} = 877.20367$	957
Total	287	3155	3442

Expected cell counts

The following table gives the expected cell count.

	Delayed	On-time	Total
American	$\frac{955 \times 287}{3442} = 79.6296$	$\frac{955 \times 3155}{3442} = 875.3704$	955
Southwest	$\frac{1530 \times 287}{3442} = 127.5741$	$\frac{1530 \times 3155}{3442} = 1402.4259$	1530
United	$\frac{957 \times 287}{3442} = 79.7963$	$\frac{957 \times 3155}{3442} = 877.20367$	957
Total	287	3155	3442

Expected cell counts

The following table gives the expected cell count.

	Delayed	On-time	Total
American	$\frac{955 \times 287}{3442} = 79.6296$	$\frac{955 \times 3155}{3442} = 875.3704$	955
Southwest	$\frac{1530 \times 287}{3442} = 127.5741$	$\frac{1530 \times 3155}{3442} = 1402.4259$	1530
United	$\frac{957 \times 287}{3442} = 79.7963$	$\frac{957 \times 3155}{3442} = 877.20367$	957
Total	287	3155	3442

Expected cell counts

The following table gives the expected cell count.

	Delayed	On-time	Total
American	$\frac{955 \times 287}{3442} = 79.6296$	$\frac{955 \times 3155}{3442} = 875.3704$	955
Southwest	$\frac{1530 \times 287}{3442} = 127.5741$	$\frac{1530 \times 3155}{3442} = 1402.4259$	1530
United	$\frac{957 \times 287}{3442} = 79.7963$	$\frac{957 \times 3155}{3442} = 877.20367$	957
Total	287	3155	3442

Significance Test of Two-Way Table Example

1. Assumptions: SRS, All of the expected cell counts are greater than 5.
2. Hypothesis:

$$H_0 : \text{Airline and on-time performance are independent.}$$
$$H_A : \text{On-time performance depends on airline.}$$

3. Test Statistic

The following table gives us the chi-square contribution for each cell, $\frac{(O-E)^2}{E}$.

	Delayed	On-time
American	$\frac{(112-79.6296)^2}{79.6296} = 13.159$	$\frac{(843-875.3704)^2}{875.3704} = 1.197$
Southwest	$\frac{(114-127.5741)^2}{127.5741} = 1.4443$	$\frac{(1416-1402.4259)^2}{1402.4259} = 0.1314$
United	$\frac{(61-79.7963)^2}{79.7963} = 4.428$	$\frac{(896-877.20367)^2}{877.20367} = 0.4028$

Test statistic:

$$X^2 = 13.159 + 1.197 + 1.4443 + 0.1314 + 4.428 + 0.4028 = 20.7625$$

3. Test Statistic

The following table gives us the chi-square contribution for each cell, $\frac{(O-E)^2}{E}$.

	Delayed	On-time
American	$\frac{(112-79.6296)^2}{79.6296} = 13.159$	$\frac{(843-875.3704)^2}{875.3704} = 1.197$
Southwest	$\frac{(114-127.5741)^2}{127.5741} = 1.4443$	$\frac{(1416-1402.4259)^2}{1402.4259} = 0.1314$
United	$\frac{(61-79.7963)^2}{79.7963} = 4.428$	$\frac{(896-877.20367)^2}{877.20367} = 0.4028$

Test statistic:

$$X^2 = 13.159 + 1.197 + 1.4443 + 0.1314 + 4.428 + 0.4028 = 20.7625$$

3. Test Statistic

The following table gives us the chi-square contribution for each cell, $\frac{(O-E)^2}{E}$.

	Delayed	On-time
American	$\frac{(112-79.6296)^2}{79.6296} = 13.159$	$\frac{(843-875.3704)^2}{875.3704} = 1.197$
Southwest	$\frac{(114-127.5741)^2}{127.5741} = 1.4443$	$\frac{(1416-1402.4259)^2}{1402.4259} = 0.1314$
United	$\frac{(61-79.7963)^2}{79.7963} = 4.428$	$\frac{(896-877.20367)^2}{877.20367} = 0.4028$

Test statistic:

$$X^2 = 13.159 + 1.197 + 1.4443 + 0.1314 + 4.428 + 0.4028 = 20.7625$$

3. Test Statistic

The following table gives us the chi-square contribution for each cell, $\frac{(O-E)^2}{E}$.

	Delayed	On-time
American	$\frac{(112-79.6296)^2}{79.6296} = 13.159$	$\frac{(843-875.3704)^2}{875.3704} = 1.197$
Southwest	$\frac{(114-127.5741)^2}{127.5741} = 1.4443$	$\frac{(1416-1402.4259)^2}{1402.4259} = 0.1314$
United	$\frac{(61-79.7963)^2}{79.7963} = 4.428$	$\frac{(896-877.20367)^2}{877.20367} = 0.4028$

Test statistic:

$$X^2 = 13.159 + 1.197 + 1.4443 + 0.1314 + 4.428 + 0.4028 = 20.7625$$

3. Test Statistic

The following table gives us the chi-square contribution for each cell, $\frac{(O-E)^2}{E}$.

	Delayed	On-time
American	$\frac{(112-79.6296)^2}{79.6296} = 13.159$	$\frac{(843-875.3704)^2}{875.3704} = 1.197$
Southwest	$\frac{(114-127.5741)^2}{127.5741} = 1.4443$	$\frac{(1416-1402.4259)^2}{1402.4259} = 0.1314$
United	$\frac{(61-79.7963)^2}{79.7963} = 4.428$	$\frac{(896-877.20367)^2}{877.20367} = 0.4028$

Test statistic:

$$X^2 = 13.159 + 1.197 + 1.4443 + 0.1314 + 4.428 + 0.4028 = 20.7625$$

3. Test Statistic

The following table gives us the chi-square contribution for each cell, $\frac{(O-E)^2}{E}$.

	Delayed	On-time
American	$\frac{(112-79.6296)^2}{79.6296} = 13.159$	$\frac{(843-875.3704)^2}{875.3704} = 1.197$
Southwest	$\frac{(114-127.5741)^2}{127.5741} = 1.4443$	$\frac{(1416-1402.4259)^2}{1402.4259} = 0.1314$
United	$\frac{(61-79.7963)^2}{79.7963} = 4.428$	$\frac{(896-877.20367)^2}{877.20367} = 0.4028$

Test statistic:

$$X^2 = 13.159 + 1.197 + 1.4443 + 0.1314 + 4.428 + 0.4028 = 20.7625$$

3. Test Statistic

The following table gives us the chi-square contribution for each cell, $\frac{(O-E)^2}{E}$.

	Delayed	On-time
American	$\frac{(112-79.6296)^2}{79.6296} = 13.159$	$\frac{(843-875.3704)^2}{875.3704} = 1.197$
Southwest	$\frac{(114-127.5741)^2}{127.5741} = 1.4443$	$\frac{(1416-1402.4259)^2}{1402.4259} = 0.1314$
United	$\frac{(61-79.7963)^2}{79.7963} = 4.428$	$\frac{(896-877.20367)^2}{877.20367} = 0.4028$

Test statistic:

$$X^2 = 13.159 + 1.197 + 1.4443 + 0.1314 + 4.428 + 0.4028 = 20.7625$$

3. Test Statistic

The following table gives us the chi-square contribution for each cell, $\frac{(O-E)^2}{E}$.

	Delayed	On-time
American	$\frac{(112-79.6296)^2}{79.6296} = 13.159$	$\frac{(843-875.3704)^2}{875.3704} = 1.197$
Southwest	$\frac{(114-127.5741)^2}{127.5741} = 1.4443$	$\frac{(1416-1402.4259)^2}{1402.4259} = 0.1314$
United	$\frac{(61-79.7963)^2}{79.7963} = 4.428$	$\frac{(896-877.20367)^2}{877.20367} = 0.4028$

Test statistic:

$$X^2 = 13.159 + 1.197 + 1.4443 + 0.1314 + 4.428 + 0.4028 = 20.7625$$

4. P-value

- The P-value for the chi-square test is $P(\chi^2 \geq X^2)$. With $df = (r-1)(c-1)$ where r = # of rows and c = # of columns.
- In our airline example r = 3, c = 2, df = (3 - 1)(2 -1) = 2.
- For our airline example, P-value = $P(\chi^2 \geq 20.7625) = 1 - pchisq(20.7625, 2) = 0.000031$

4. P-value

- The *P*-value for the chi-square test is $P(\chi^2 \geq X^2)$. With $df = (r - 1)(c - 1)$ where r = # of rows and c = # of columns.
- In our airline example r = 3, c = 2, df = (3 - 1)(2 -1) = 2.
- For our airline example, *P*-value = $P(\chi^2 \geq 20.7625) = 1 - pchisq(20.7625, 2) = 0.000031$

4. P-value

- The P-value for the chi-square test is $P(\chi^2 \geq X^2)$. With $df = (r-1)(c-1)$ where r = # of rows and c = # of columns.
- In our airline example r = 3, c = 2, df = (3 - 1)(2 -1) = 2.
- For our airline example, P-value = $P(\chi^2 \geq 20.7625) = 1 - pchisq(20.7625, 2) = 0.000031$

5. Decision

- **Reject** H_0 if the P-value $\leq \alpha$.
- **Fail** to reject H_0 if the $P - \text{value} > \alpha$.
- In our airplane example, $P - \text{value} < 0.0001$ so we **reject** the null hypothesis.

5. Decision

- **Reject** H_0 if the P-value $\leq \alpha$.
- **Fail** to reject H_0 if the $P - \text{value} > \alpha$.
- In our airplane example, $P - \text{value} < 0.0001$ so we **reject** the null hypothesis.

6. Conclusion

- If H_0 is **rejected** then there is a dependence between the row variable and the column variable.
- If H_0 is not rejected then there is no association.
- In our airplane example, we **reject** the null hypothesis. Thus we conclude that on-time status **depends** on airline.

6. Conclusion

- If H_0 is **rejected** then there is a dependence between the row variable and the column variable.
- If H_0 is not rejected then there is no association.
- In our airplane example, we **reject** the null hypothesis. Thus we conclude that on-time status **depends** on airline.

Chi-square Test Using R

1. Input the data as a matrix.
2. R-code: chisq.test(matrix name,correction=FALSE)

```
> airline<-matrix(c(112,114,61,843,1416,896),nrow=3,ncol=2)
> chisq.test(airline,correct = FALSE)

Pearson's Chi-squared test

data:  airline
X-squared = 20.762, df = 2, p-value =3.102e-05
```

Understanding Dependence

- By itself, the chi-square test determines only whether the data provide evidence of a relationship between the two variables. If the result is significant, one can go on to identify the source of that relationship by finding the cells of the table that contribute most to the χ^2 value (i.e. those cells with the biggest discrepancy between the observed and expected counts) and by noting whether the observed count falls above or below the observed count in those cells.

- To get these "Chi-square contribution" values in R use `residuals(chisq.test(matrix,correction=FALSE))^2`.

Eating Out

A survey was conducted in five countries. The following table is based on 1,000 respondents in each country that said they eat out once a week or more (yes) or not (no).

	Country				
Eat out	Germany	France	UK	Greece	US
Yes	100	120	280	390	570
No	900	880	720	610	430

At the 0.05 level of significance, determine whether there is a significant difference in the proportion of people who eat out at least once a week in the various countries.

R Output

```
> eat<-matrix(c(100,900,120,880,280,720,390,610,570,430),nrow = 2
,ncol = 5)
> eat
[,1] [,2] [,3] [,4] [,5]
[1,]  100  120  280  390  570
[2,]  900  880  720  610  430
> chisq.test(eat,correct = FALSE)

Pearson's Chi-squared test

data:  eat
X-squared = 742.4, df = 4, p-value < 2.2e-16

> residuals(chisq.test(eat,correct = FALSE))^2
[,1]        [,2]        [,3]       [,4]       [,5]
[1,] 126.2466 101.31507 0.4931507 32.89041 264.6712
[2,]  52.0678  41.78531 0.2033898 13.56497 109.1582
```

Review Questions

Use the data below to determine if there is sufficient evidence to conclude that an association exists between car color and the likelihood of being in an accident.

	Red	Blue	White
Car has been in accident	28	33	36
Car has not been in accident	23	22	30

R output:

```
X-squared = 0.42871, df = 2, p-value = 0.8071
```

1. Give the null hypothesis for this test.
   a) There is no association between car color and the likelihood of being in an accident.
   b) There is an association between car color and the likelihood of being in an accident.
   c) Car color and likelihood of being in an accident are equal.
   d) There are more white cars that are in an accident.

Review Question

R output:

```
X-squared = 0.42871, df = 2, p-value = 0.8071
```

2. Give the decision of the test, use $\alpha = 0.05$.
   a) Reject H_0.
   b) Fail to reject H_0.
   c) Accept H_0.
   d) H_0 is false.

3. Which of the following is a valid conclusion for this test?
   a) We have sufficient evidence of an association between car color and the likelihood of being in an accident at the 5% level.
   b) There is insufficient evidence of an association between car color and the likelihood of being in an accident at the 5% level.
   c) White cars are in an accident more often.
   d) Red cars are less likely to be in an accident.

What to Expect on the Exam

The test has two parts

1. 14 questions multiple choice.
2. 5 free response questions.
3. See examples of problems in quizzes or tests.

Example 1

The weights of male and female students in a class are summarized in the following boxplots:

Example 1 continued

Which of the following is NOT correct?

a) The median weight of the male students is about 166 lbs.

b) The mean weight of the female students is about 120 because of symmetry.

c) The male students have less variability than the female students.

d) About 50% of the male students have weights between 150 and 185 lbs.

Example 2

Hospital records show that 12% of all patients are admitted for heart disease, 28% are admitted for cancer (oncology) treatment, and 6% receive both coronary and oncology care.

1. What is the probability that a randomly selected patient is admitted for coronary care, oncology or both? (Note that heart disease is a coronary care issue.)

2. What is the probability that a randomly selected patient is admitted for coronary care, given that they are a cancer patient?

3. Are patients that are admitted for coronary care independent of patients that are admitted for cancer?

Example 3

A random variable X has a probability distribution as follows:

X	0	1	2	3	4
$P(X)$	$2k$	$3k$	$5k$	$3k$	$4k$

1. What is the value of k?

2. What is $P(X < 2)$?

Example 4

In testing a certain kind of missile, target accuracy is measured by the average distance X (from the target) at which the missile explodes. The distance X is measured in miles and the distribution of X is given by:

X	0	10	50	100
P(X)	$\frac{1}{14}$	$\frac{1}{7}$	$\frac{2}{7}$	$\frac{1}{2}$

Find the mean and variance for the target accuracy.

Example 4

In testing a new drug, researchers found that 5% of all patients using it will have a mild side effect. A random sample of 7 patients using the drug is selected. Find the probability that:

1. None will have this mild side effect.

2. Exactly 2 patients will have this mild side effect.

3. At least one will have this mild side effect.

4. What is the expected value and variance of the number of patients that will have this mild side effect?

Example 5

Let X be the amount of time (in hours) the wait is to get a table at a restaurant. Suppose the cdf is represented by

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{4}x^2 & 0 \leq x \leq 2 \\ 1 & x > 2 \end{cases}$$

1. Find $P(X \leq 1.5)$
2. Find $P(X \geq 1)$
3. Find $P(1 \leq X \leq 1.5)$
4. Find the density function $f(x)$.

Example 6

Let Z be the standard normal random variable. Calculate the following.

1. $P(|Z| \leq 2.4)$

2. $P(Z \leq -1.9)$

3. Find c such that $P(Z \geq c) = 0.98$

Example 7

The weights of individual bolts produced at a manufacturing plant, X, is normally distributed. If the mean weight of the bolts is 9 grams and the standard deviation is 3.2 grams, find:

1. $P(X \leq 10.5)$

2. $P(X \geq 7.1)$

3. The value of x such that $P(X \leq x) = 0.93$

Example 8

Using the information from example 7, if we randomly sample 62 weights and determine the sample mean, what is the probability that

1. The sample mean is less than 8.5 grams.

2. The sample mean is at least 10.5 grams.

Example 9

1. 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.

2. An SRS of 24 students at UH gave an average height of 6.1 feet and a standard deviation of .3 feet. Construct a 90% confidence interval for the mean height of students at UH.

3. The average height of students at UH from an 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.

Example 10

A 98% confidence interval for the mean of a population is to be constructed and must be accurate to within 0.3 unit. A preliminary sample standard deviation is 1.7. The smallest sample size n that provides the desired accuracy is

Example 11

In a hypothesis test, if the computed P-value is less than 0.001, there is very strong evidence to

a) retest with a different sample.
b) accept the null hypothesis
c) fail to reject the null hypothesis.
d) reject the null hypothesis.

Example 12

The one-sample t statistic for a test of $H_0 : \mu = 12$ vs. $H_a : \mu < 12$ based on $n = 174$ observations has the test statistic value of $t = -1.58$. What is the p-value for this test?

Example 13

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?

a) One sample t test
b) Matched pairs
c) Two sample t test
d) Two sample p test

Complete the Test

Example 14

Data for gas mileage (in mpg) for different vehicles was entered into a software package and part of the ANOVA table is shown below:

Source	DF	SS	MS
Vehicle	2	440	220.00
Error	17	318	18.71
Total	19	758	

1. Determine the value of the test statistic F to complete the table.

2. Determine the p-value.

Example 15

The community hospital is studying its distribution of patients. A random sample of 330 patients presently in the hospital gave the following information:

Type of Patient	Old Rate	Number of Occurrences
Maternity Ward	20%	77
Cardiac Ward	32%	95
Burn Ward	10%	29
Children's Ward	15%	53
All Other Wards	23%	76

Test the claim at the 5% significance level that the distribution of patients in these wards has not changed.

Example 15 Work

Example 16

The following two-way table describes the preferences in movies and fast food restaurants for a random sample of 100 people.

	McDonald's	Taco Bell	Wendy's
Iron Man	20	12	8
Dispicable Me	12	7	9
Harry Potter	6	14	12

1. What percent of the Dispicable Me lovers also like McDonald's?

2. What percent likes Harry Poter if they also like Wendy's?

Example 17

Below is the computer output for the appraised value (in thousands of dollars) and number of rooms for 20 houses in East Meadow, New York.

Predictor	Coef	Stdev	t-ratio
Constant	74.80	19.04	3.93
Rooms	19.718	2.631	7.49

$S = 29.05$ R-sq = 43.8% R-sq (adj) = 43.0%

1. What is the regression equation?

2. Predict the price of a 10 room house (in thousands of dollars).

3. Calculate the 95% confidence interval of the slope of the regression line for all homes.

Use the information provided to test whether there is a significant relationship between the price of a house and the number of rooms at the 5% level.

Example 18

The following data are for intelligence-test (IT) scores, grade-point averages (GPA), and reading rates (RR) of at-risk students.

IT	295	152	214	171	131	178	225	141	116	173
GPA	2.4	.6	.2	0	1	.6	1	.4	0	2.6
RR	41	18	45	29	28	38	25	26	22	37

a) Calculate the line of best fit that predicts the GPA on the basis of IT scores.

b) Calculate the line of best fit that predicts the GPA on the basis of RR scores.

c) Which of the two lines calculated in parts a and b best fits the data?

What You Need an What is Provided

- Provided
  - Online calculator; it will be a link you see in the exam.
  - R; it will be a link you see in the exam.
  - Formula sheet and z, t, chi-square tables.
- Can bring
  - Pencil; you will need something to write with for the free response questions.
  - Your Cougar Card.

Questions?