



Bookmarks

▶ Important Pre-Course Survey

▶ Contact Us

▶ How To Navigate the Course

▶ Discussion Board

▶ Office Hours

▶ Week 0: Introduction to Data (Optional Review)

▶ Week 1: Sampling

▶ Week 2: Hypothesis Testing (One Group Means)

▶ Week 3: Hypothesis Testing (Two Group Means)

▼ Week 4: Hypothesis Testing (Categorical Data)

Readings

Week 4: Hypothesis Testing (Categorical Data) > Pre-Lab > Prepare for the Analysis



Bookmark

Reflect on the Question

Analyze the Data

Draw Conclusions

Primary Research Questions

1. Are there an equal number of male and female performers on Austin City Limits?
2. Are male performers just as likely to have had a Top 10 hit as female performers?

Breakdown Your Analysis

Let's break this analysis into its required steps:

Goodness of Fit Test:

1. Make a table of counts for gender.
2. Create a vector of the expected proportions.
3. Check the expected counts assumption.
4. Run the chi square test.
5. Interpret the chi square statistic and p-value.


Test of Independence:

1. Create a two-way table for gender and Top 10 hits.
2. Check the expected counts assumption.
3. Run the chi square test.
4. Interpret the chi square statistic and p-value.


Here is the code you will use:

Question 1 (Goodness of Fit)

```
# Create a table of counts for Gender
gender_tab <- table(ac1$Gender)
gender_tab
```


Reading Check due
May 03, 2016 at 17:00
UTC 

Lecture Videos


Comprehension Check
due May 03, 2016 at
17:00 UTC 

R Tutorial Videos


Pre-Lab

Pre-Lab due May 03,
2016 at 17:00 UTC 

Lab

Lab due May 03, 2016
at 17:00 UTC 

Problem Set

Problem Set due May
03, 2016 at 17:00 UTC 

Create vector of expected proportions

```
ExpGender <- c(.50, .50)
```

Check expected counts assumption

```
chisq.test(gender_tab, p=ExpGender)$expected
```

Run goodness of fit

```
chisq.test(gender_tab, p=ExpGender)
```

Question 2 (Test of Independence)

Create two-way table

```
gender_top10 <- table(acl$Gender, acl$BB.wk.top10)
```

```
gender_top10
```

Generate expected counts

```
chisq.test(gender_top10, correct=FALSE)$expected
```


Run test of independence

```
chisq.test(gender_top10, correct=FALSE)
```

(1/1 point)

1. If we wanted to test the hypothesis that the performers were 30% female and 70% male, what would the code look like? (Note that categorical values are referenced in alphabetical order).

☐ ExpGender <- c(.70, .30)

☒ ExpGender <- c(.30,.70) 

☐ ExpGender <- c(.50, .50)

[Click here for a video explanation of how to answer this question.](#)

You have used 1 of 1 submissions

(1 point possible)

2. Suppose the following values were returned for the "check expected counts" assumption in our goodness of fit test. Would the assumption be violated?

F M

3 29

☐ No, because the total number of expected counts is greater than 5.

☒ Yes, because there are fewer females than males expected, and the test proportion is 50/50. ❌

☐ Yes, because there are fewer than 5 expected Females. ✔️

[Click here for a video explanation of how to answer this question.](#)

You have used 1 of 1 submissions

(1/1 point)

3. Which line of code is **not** necessary for a test of independence because there is no particular distribution model being tested?

☐ `gender_top10 <-table(acl$Gender, acl$BB.wk.top10)`

☒ `ExpGender_top10 <- c(.25, .25, .25, .25)` ✔️

☐ `chisq.test(gender_top10, correct=FALSE)`

[Click here for a video explanation of how to answer this question.](#)

You have used 1 of 1 submissions

(1/1 point)

4. How many degrees of freedom should there be for our test of independence? Remember, performers have either had (or not had) a Top 10 hit.

☒ 1 ✓

☐ 2

☐ 3

[Click here for a video explanation of how to answer this question.](#)

You have used 1 of 1 submissions

(1/1 point)

Suppose we wanted to test whether there was an even distribution among the seasons. What caused the following error below? (You may want to examine the dataset in R for help.)

```
acl <- AustinCityLimits
season_counts <- table(acl$Season)
ExpSeason <- c(1/4, 1/4, 1/4, 1/4)
chisq.test(season_counts, p=ExpSeason)
```

Error in chisq.test(season_counts, p = ExpSeason) : 'x' and 'p' must have the same number of elements

☐ The probabilities in line 3 need to be expressed as decimals instead of fractions.

☐ We ran the wrong Chi Square test.

☒ There are not 4 seasons in our dataset, although line 3 suggests that there are. ✓

☐ The variable "Season" is not capitalized in our dataset.

[Click here for a video explanation of how to answer this question.](#)

You have used 1 of 1 submissions

© All Rights Reserved



© edX Inc. All rights reserved except where noted. EdX, Open edX and the edX and Open EdX logos are registered trademarks or trademarks of edX Inc.

POWERED BY
OPENedX

