Chi-Square Tests

Data Analysis for Psychology in R 1

Semester 2 Week 9

Dr Emma Waterston

Department of Psychology
The University of Edinburgh

Course Overview

Exploratory Data Analysis	Research design and data
	Describing categorical data
	Describing continuous data
	Describing relationships
	Functions
Probability	Probability theory
	Probability rules
	Random variables (discrete)
	Random variables (continuous)
	Sampling

Foundations of inference	Confidence intervals
	Hypothesis testing (p-values)
	Hypothesis testing (critical values)
	Hypothesis testing and confidence intervals
	Errors, power, effect size, assumptions
Common hypothesis tests	One sample t-test
	Independent samples t-test
	Paired samples t-test
	Chi-square tests
	Correlation

Learning Objectives

- Understand the difference between χ^2 goodness-of-fit and χ^2 test of independence
- Understand how to perform a χ^2 goodness-of-fit and interpret results
- Understand how to perform a χ^2 test of independence and interpret results

Introduction to χ^2

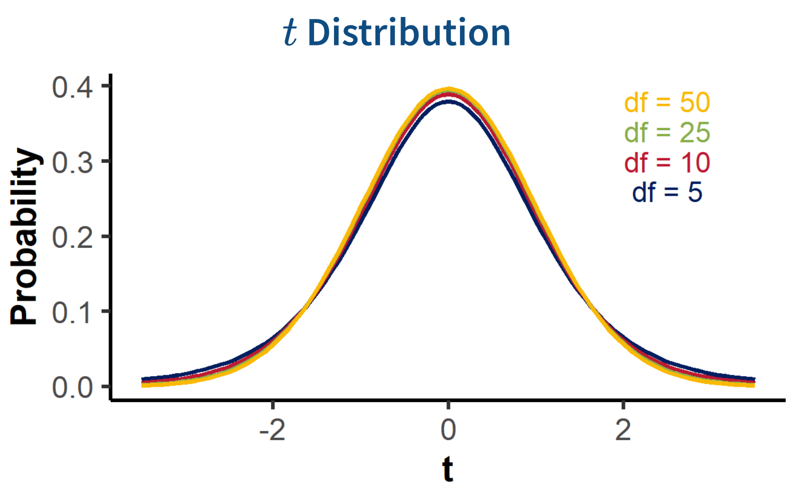
Moving on From t-tests...

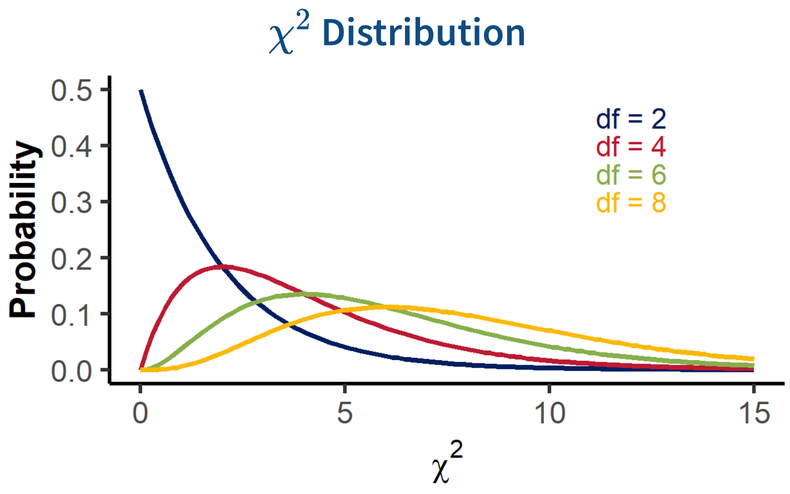
- t-tests have allowed you to make comparisons using *continuous* data:
  - One continuous variable against a single value (one sample t-test)
  - A continuous outcome variable from two separate groups (independent samples t-test)
  - A continuous outcome variable from one group at two time points (paired samples t-test)
- You may instead want to test whether data are distributed across *categories* in the way that you would expect:
  - Is your sample distributed equally across levels of education?
  - Is smoking (Y/N) associated with cardiovascular disease (Y/N)?
- In this case, you will will need a test that checks whether data are grouped according to your expectations
  - χ^2-tests are used to compare **frequencies** across categories in your data

χ^2-tests vs t-tests

- Process similar to a t-test:
    1. Compute a test statistic
    2. Locate the test statistic on a distribution that reflects the probability of each test statistic value, given that H_0 is true
    3. If the probability associated with your test statistic is small enough, your results are considered significant

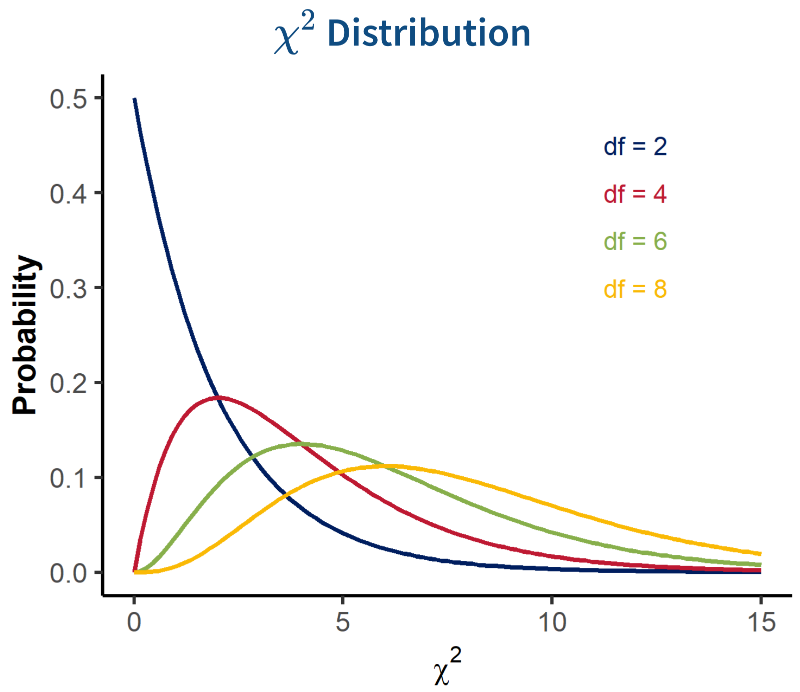
- Distribution
    - Like the t-distribution, the shape of the distribution depends on the degrees of freedom
    - Unlike the t-distribution, *df* in a χ^2 test isn't computed using sample size, but the number of groups within your data





χ^2 Distribution

- As the number of comparison groups increases, the distribution curve flattens
  - Larger χ^2 values become more probable
  - A wider range of χ^2 values become more likely
- The χ^2 distribution begins at 0
  - Categorical variables don't have direction
  - p-value is computed only in one direction (right-tail) as the Probability of observing a χ^2 statistic as big or bigger than the one obtained
  - We can investigate this further by looking at the χ^2 formula



Data Requirements & Assumptions of χ^2 Tests

- Data Requirements
  - Variables should be measured at an ordinal or nominal level (i.e., categorical data)

- Assumptions
  - Expected counts ≥ 5
  - Observations are independent
    - Each observation appears only in a single cell

Types of χ^2 Tests

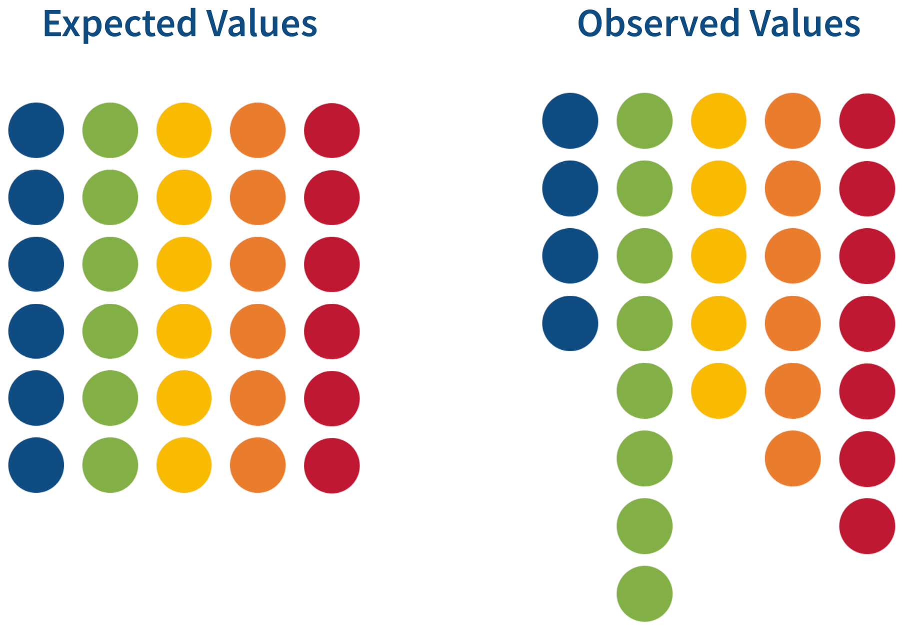
- Goodness of Fit
- Test of Independence

Questions?

χ^2 Goodness of Fit Test

χ^2 Goodness of Fit Test

- Tests whether the proportions / relative frequencies you actually have are consistent with the expected proportions / relative frequencies
- Looks at the distribution of data across a single category
- **Hypotheses:**
  - $H_0 : p_1 = p_{1,0},\ p_2 = p_{2,0},\ \ldots,\ p_k = p_{k,0}$
  - H_1 : Some $p_i \neq p_{i,0}$



χ^2 Goodness of Fit Test

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

- where:
  - Σ = sum up
  - $\sum_{i=1}^{k}$: Sum all values from levels 1 through k
  - E = Expected Cases
    - The values that you expect, given H_0 is true
  - O = Observed Cases
    - The values you actually have
  - i : Current level

χ^2 Goodness of Fit Test Example

- A new flower shop is trying to decide which days of the week they will be open
- They want to know whether order number is consistent across days of the week
- They count the total number of orders they take each day of the week over the course of a month

Data

```
## # A tibble: 7 × 2
##   Day       Orders
##   <fct>      <dbl>
## 1 Monday        54
## 2 Tuesday       39
## 3 Wednesday     44
## 4 Thursday      47
## 5 Friday        68
## 6 Saturday      72
## 7 Sunday        53
```

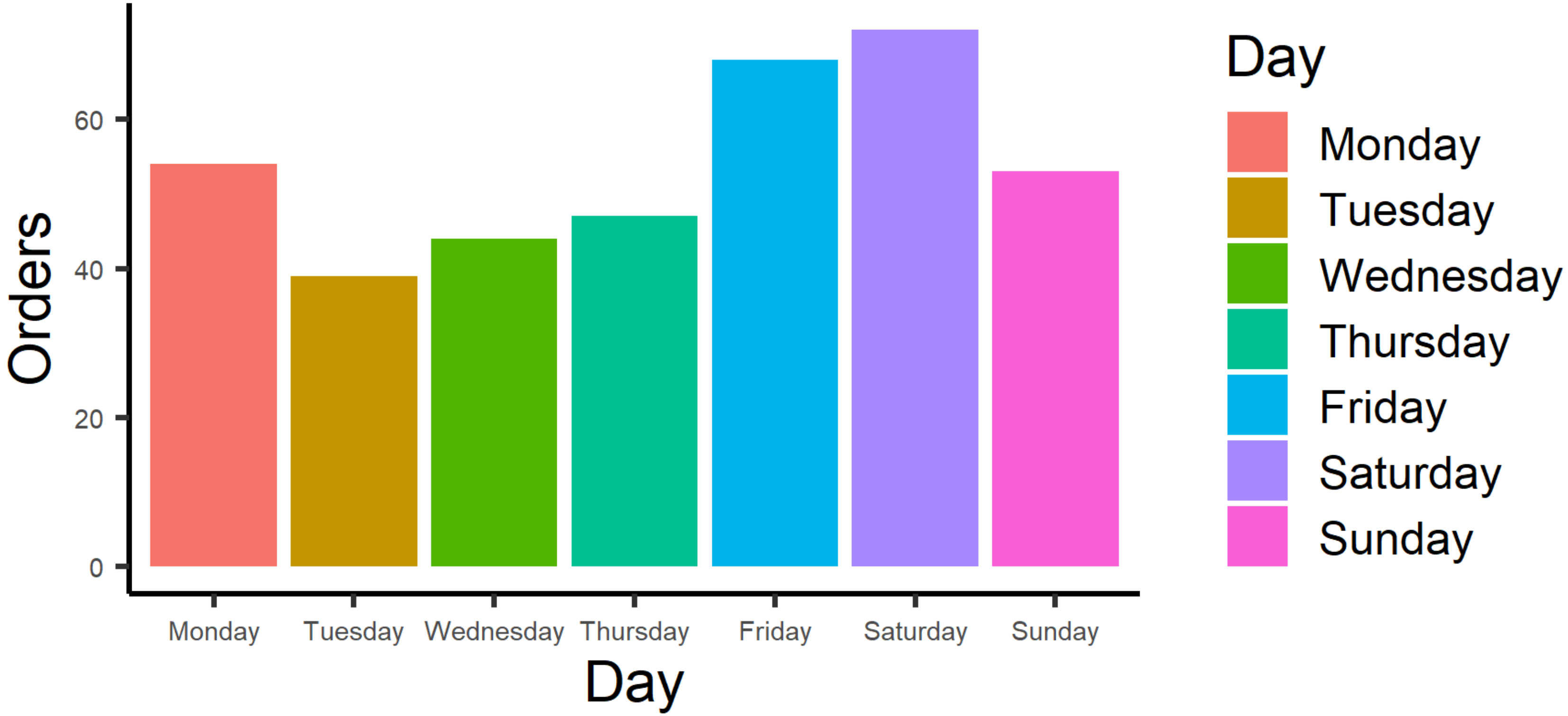
Hypotheses

- I elect to use an alpha (α) of .05
- My hypotheses are:
  - H_0: Orders will be consistent throughout the week
    - $p_{Monday} = p_{Tuesday} = \cdots = p_{Sunday}$
  - H_1: Orders will differ across the week
    - Some $p_i \neq p_{i0}$

Day	Orders
Monday	54
Tuesday	39
Wednesday	44
Thursday	47
Friday	68
Saturday	72
Sunday	53

Visualisation

```
ggplot(data = flowerDat, aes(Day, Orders, fill = Day)) +
  geom_col()
```



Performing a χ^2 Goodness of Fit Test

Compute the test statistic

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

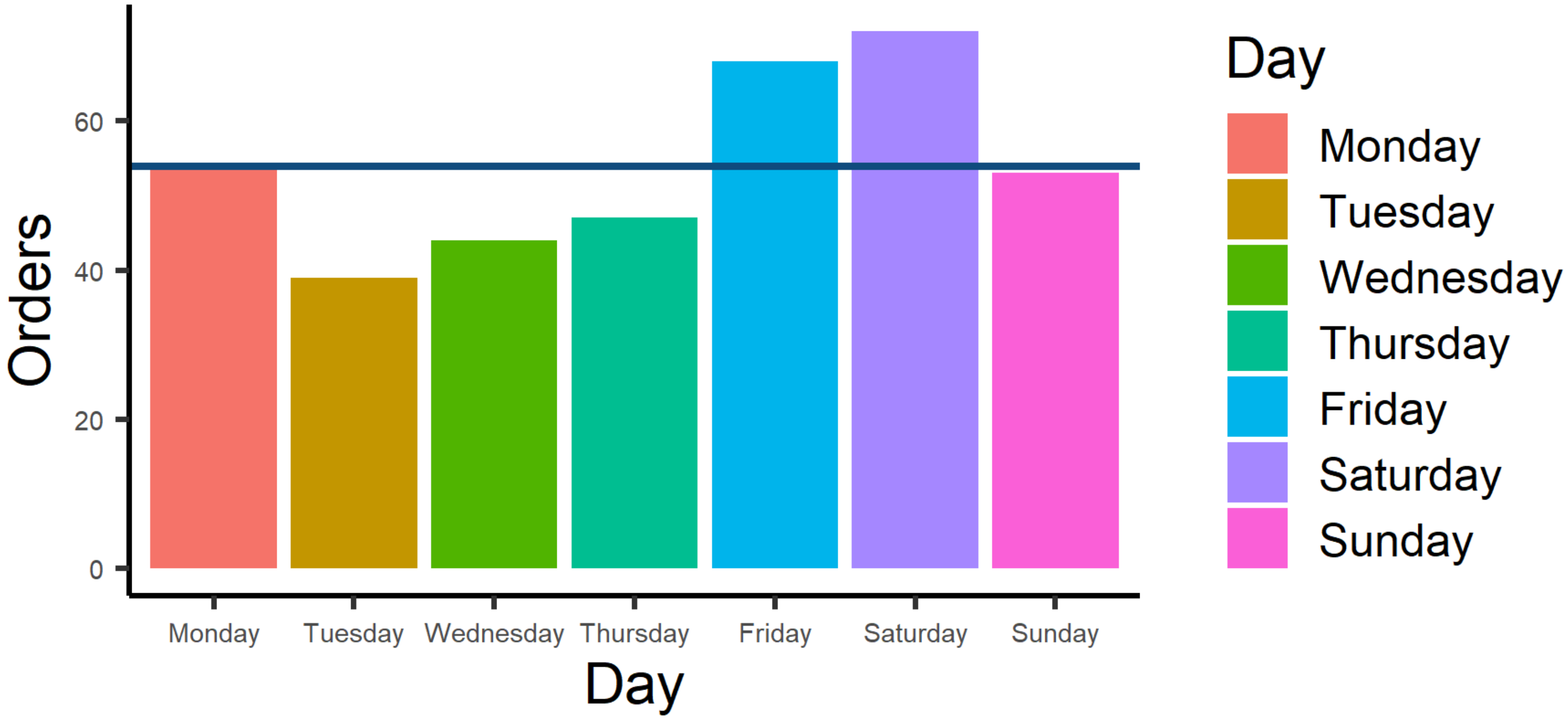
- $E_i = n \cdot p_i$
- where:
    - n = sample size
    - p = the hypothesized population proportion for the category under the null hypothesis
- In this example, we expect each level to be approximately equal, so the expected proportion will be the same across levels:

```
n <- sum(flowerDat$Orders)
p <- (1/length(levels(flowerDat$Day))) # i.e., 1/7

E <- n * p
round(E, digits = 2)
```

```
## [1] 53.86
```

Visualisation



Performing a χ^2 Goodness of Fit Test

Compute the test statistic

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Day	Orders	Expected
Monday	54	53.86
Tuesday	39	53.86
Wednesday	44	53.86
Thursday	47	53.86
Friday	68	53.86
Saturday	72	53.86
Sunday	53	53.86

Performing a χ^2 Goodness of Fit Test

Compute the test statistic

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Day	Orders	Expected	Difference
Monday	54	53.86	0.14
Tuesday	39	53.86	-14.86
Wednesday	44	53.86	-9.86
Thursday	47	53.86	-6.86
Friday	68	53.86	14.14
Saturday	72	53.86	18.14
Sunday	53	53.86	-0.86

Performing a χ^2 Goodness of Fit Test

Compute the test statistic

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Day	Orders	Expected	Difference	Squared
Monday	54	53.86	0.14	0.02
Tuesday	39	53.86	-14.86	220.73
Wednesday	44	53.86	-9.86	97.16
Thursday	47	53.86	-6.86	47.02
Friday	68	53.86	14.14	200.02
Saturday	72	53.86	18.14	329.16
Sunday	53	53.86	-0.86	0.73

Performing a χ^2 Goodness of Fit Test

Compute the test statistic

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Day	Orders	Expected	Difference	Squared	SqbyExp
Monday	54	53.86	0.14	0.02	0.00
Tuesday	39	53.86	-14.86	220.73	4.10
Wednesday	44	53.86	-9.86	97.16	1.80
Thursday	47	53.86	-6.86	47.02	0.87
Friday	68	53.86	14.14	200.02	3.71
Saturday	72	53.86	18.14	329.16	6.11
Sunday	53	53.86	-0.86	0.73	0.01

Performing a χ^2 Goodness of Fit Test

Compute the test statistic

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = 16.62$$

Day	Orders	Expected	Difference	Squared	SqbyExp
Monday	54	53.86	0.14	0.02	0.00
Tuesday	39	53.86	-14.86	220.73	4.10
Wednesday	44	53.86	-9.86	97.16	1.80
Thursday	47	53.86	-6.86	47.02	0.87
Friday	68	53.86	14.14	200.02	3.71
Saturday	72	53.86	18.14	329.16	6.11
Sunday	53	53.86	-0.86	0.73	0.01

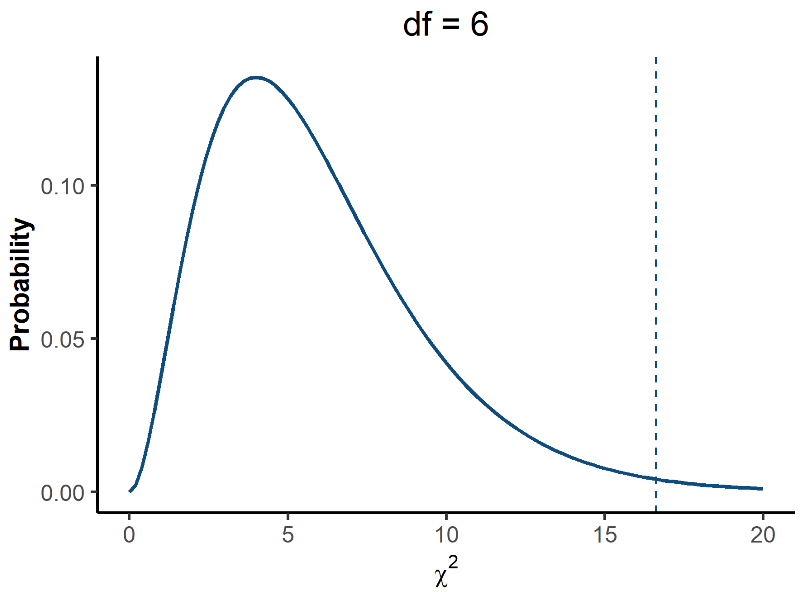
Performing a χ^2 Goodness of Fit Test

Find the test statistic on the distribution

$\chi^2 = 16.62$

$df = k - 1$

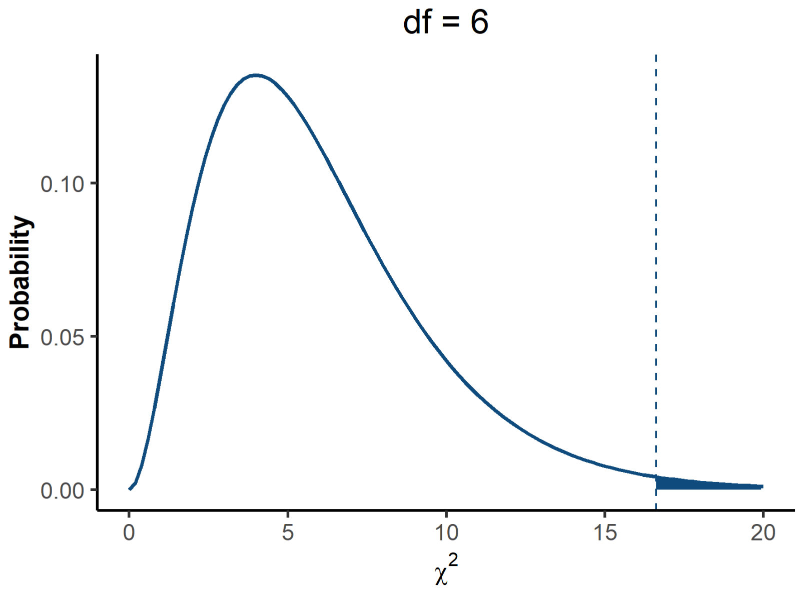
- where k = number of levels within categorical variable
- so, in our example:
    - k = number of days in the week
    - $df \ = \ (7 - 1) \ = \ 6$



Performing a χ^2 Goodness of Fit Test

Compute the probability of obtaining a χ^2 statistic at least as extreme as the observed one, if H_0 is true

- What proportion of the plot falls in the shaded area?



Performing a χ^2 Goodness of Fit Test

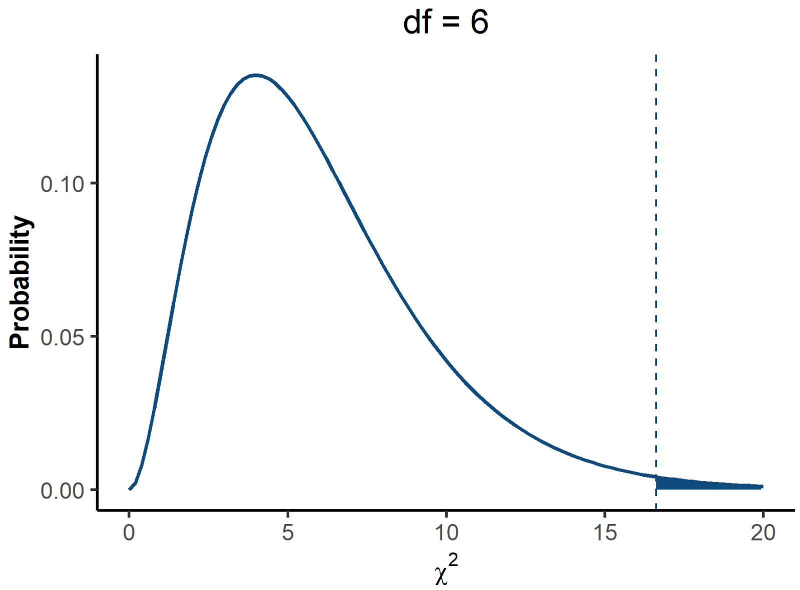
Compute the probability of obtaining a χ^2 statistic at least as extreme as the observed one, if H_0 is true

- What proportion of the plot falls in the shaded area?

```
pchisq(x2_stat_gof,
       df = 6,
       lower.tail = FALSE)
```

```
## [1] 0.01080571
```

- The probability that we would have a χ^2 value as extreme as 16.62 if H_0 is true is only 0.011



Performing a χ^2 Goodness of Fit Test in R

```r
#Option 1
observed <- c(54, 39, 44, 47, 68, 72, 53)
expected <- c(1/7,1/7,1/7,1/7,1/7,1/7,1/7)
GOFtest <- chisq.test(x = observed, p = expecte
GOFtest
```

```
##
##      Chi-squared test for given probabilities
##
## data:   observed
## X-squared = 16.615, df = 6, p-value = 0.01081
```

- where:
  - x: A numerical vector of observed frequencies
  - p: A numerical vector of expected proportions

```r
#Option 2
GOFtest <- chisq.test(flowerDat$Orders)
GOFtest
```

```
##
##      Chi-squared test for given probabilities
##
## data:   flowerDat$Orders
## X-squared = 16.615, df = 6, p-value = 0.01081
```

Exploring our Results Further

- If our results are significant, we are likely interested in knowing which levels within our category had the biggest differences
- We can get this information by looking at the Pearson residuals (AKA, *standardized residuals*)
  - $\frac{O_i - E_i}{\sqrt{E_i}}$

In R

```
GOFtest$residuals
```

```
## [1]  0.01946616 -2.02448072 -1.34316509
## [4] -0.93437571  1.92714991  2.47220241
## [7] -0.11679696
```

By Hand

- Need to calculate separately for each level
- *Example of number of flowers sold on a Monday:*
  $$\frac{54 - 53.86}{\sqrt{53.86}} = 0.019$$

Exploring our Results Further

- Positive residuals indicate that the observed frequency of the corresponding level is higher than the expected frequency
- Negative residuals indicate that the observed frequency of the corresponding level is lower than the expected frequency
- More extreme residuals indicate that the values are contributing more strongly to the results
  - Values $\leq$ -2 indicate the observed frequency of that level is **much lower** than expected
  - Values $\geq$ 2 indicate the observed frequency of that level is **much higher** than expected

Day	Orders	Residuals
Monday	54	0.02
Tuesday	39	-2.02
Wednesday	44	-1.34
Thursday	47	-0.93
Friday	68	1.93
Saturday	72	2.47
Sunday	53	-0.12

Drawing Conclusions

If you owned the flower shop, which two days would you choose to close each week?

Day	Orders	Residuals
Monday	54	0.02
Tuesday	39	-2.02
Wednesday	44	-1.34
Thursday	47	-0.93
Friday	68	1.93
Saturday	72	2.47
Sunday	53	-0.12

Write Up

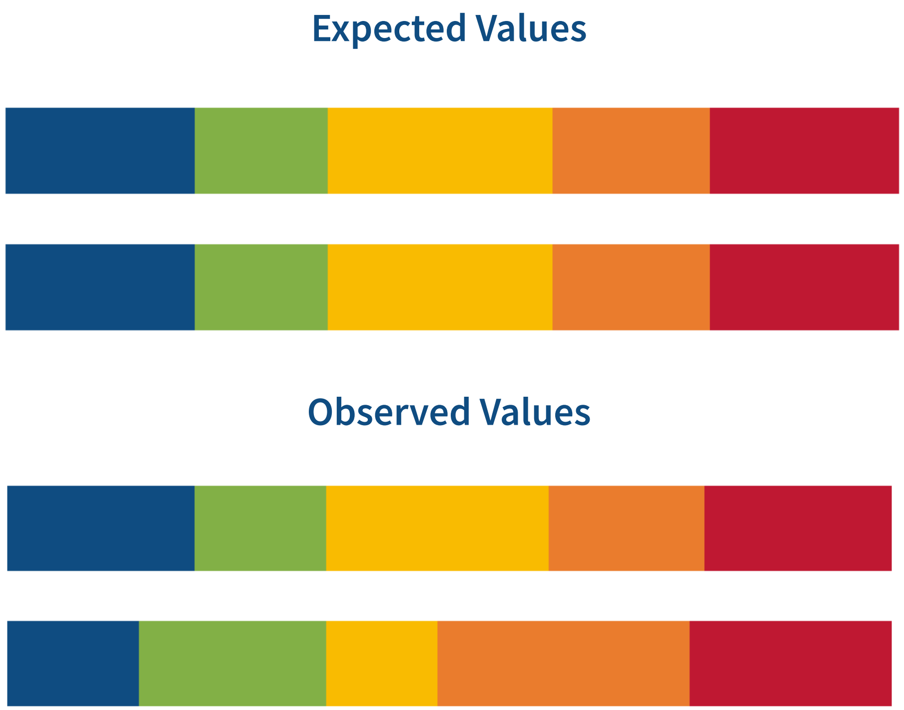
A χ^2 Goodness of Fit test was conducted in order to determine whether the proportion of flower orders was equal across each day of the week. The goodness of fit test was significant $(\chi^2(6, n = 377) = 16.62, p = .011)$, and thus, with $\alpha = .05$, we would reject the null hypothesis as the proportion of flower orders differed across the days of the week.

Questions?

χ^2 Test of Independence

χ^2 Test of Independence

- Checks whether two categorical variables from a single population are independent of each other
- Specifically, tests whether membership in Variable 1 is dependent upon membership in Variable 2
- **Hypotheses:**
  - H_0 : Variables A and B are independent
  - H_1 : There is an association between Variable A and Variable B



χ^2 Test of Independence

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

- where:
  - Σ = sum up
  - E = Expected cases
  - O = Observed cases
  - i : current level within Variable A
  - j : current level within Variable B
  - r : total levels within Variable A
  - c : total levels within Variable B

Example

- The flower shop is trying to decide on their flower stock
- They want to know whether the flower type that sells the best depends on the season
- **Hypotheses:**
  - H_0: Flower orders will be independent of season
  - H_1: Flower orders will be dependent on season

Data

First 6 rows:

```
## # A tibble: 6 × 2
##   Season Flowers
##   <fct>  <fct>
## 1 Spring Roses
## 2 Spring Roses
## 3 Spring Roses
## 4 Spring Roses
## 5 Spring Roses
## 6 Spring Roses
```

Create a contingency table:

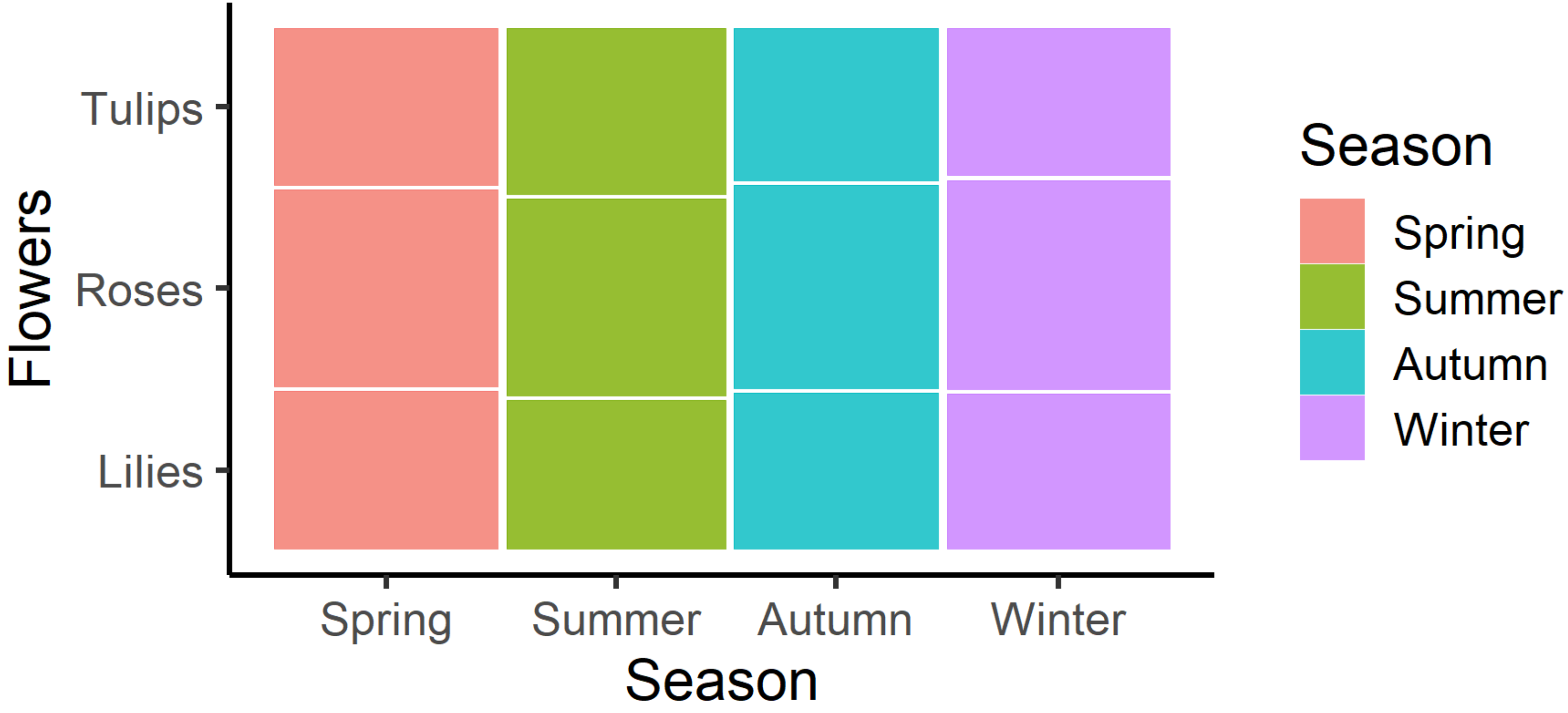
```
#Option 1
xtabs(~ Season + Flowers, data = seasonDat)

#Option 2
table(seasonDat$Season, seasonDat$Flowers)
```

```
##
##          Lilies Roses Tulips
##   Spring    186   232    185
##   Summer    172   228    192
##   Autumn    168   219    164
##   Winter    183   246    173
```

Visualisation

```
library(ggmosaic)
ggplot(data = seasonDat) +
  geom_mosaic(aes(x = product(Flowers, Season), fill = Season))
```



Performing a χ^2 Test of Independence

Compute the test statistic

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

- $E_{ij} = \frac{R_i \cdot C_j}{n}$
- In this example, we expect the orders to be distributed evenly across season and flower type

Performing a χ^2 Test of Independence

Compute the test statistic

- $E_{ij} = \frac{R_i \cdot C_j}{n}$

	Lilies	Roses	Tulips	Sum
Spring	186	232	185	603
Summer	172	228	192	592
Autumn	168	219	164	551
Winter	183	246	173	602
Sum	709	925	714	2348

Season	Lilies	Roses	Tulips
Spring	$\frac{(603 * 709)}{2348}$	$\frac{(603 * 925)}{2348}$	$\frac{(603 * 714)}{2348}$
Summer	$\frac{(592 * 709)}{2348}$	$\frac{(592 * 925)}{2348}$	$\frac{(592 * 714)}{2348}$
Autumn	$\frac{(551 * 709)}{2348}$	$\frac{(551 * 925)}{2348}$	$\frac{(551 * 714)}{2348}$
Winter	$\frac{(602 * 709)}{2348}$	$\frac{(602 * 925)}{2348}$	$\frac{(602 * 714)}{2348}$

Performing a χ^2 Test of Independence

Compute the test statistic

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Observed Counts

	Lilies	Roses	Tulips
Spring	186	232	185
Summer	172	228	192
Autumn	168	219	164
Winter	183	246	173

Expected Counts

Seasons	Lilies	Roses	Tulips
Spring	182.08	237.55	183.37
Summer	178.76	233.22	180.02
Autumn	166.38	217.07	167.55
Winter	181.78	237.16	183.06

Performing a χ^2 Test of Independence

Compute the test statistic

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Observed Counts

	Lilies	Roses	Tulips
Spring	186	232	185
Summer	172	228	192

Expected Counts

Seasons	Lilies	Roses	Tulips
Spring	182.08	237.55	183.37
Summer	178.76	233.22	180.02

Difference

Seasons	Lilies	Roses	Tulips
Spring	3.92	-5.55	1.63
Summer	-6.76	-5.22	11.98

Performing a χ^2 Test of Independence

Compute the test statistic

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Difference

Seasons	Lilies	Roses	Tulips
Spring	3.92	-5.55	1.63
Summer	-6.76	-5.22	11.98
Autumn	1.62	1.93	-3.55
Winter	1.22	8.84	-10.06

Squared

Seasons	Lilies	Roses	Tulips
Spring	15.36	30.84	2.67
Summer	45.69	27.25	143.51
Autumn	2.63	3.73	12.62
Winter	1.49	78.16	101.23

Performing a χ^2 Test of Independence

Compute the test statistic

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

Squared

Seasons	Lilies	Roses	Tulips
Spring	15.36	30.84	2.67
Summer	45.69	27.25	143.51

Expected

Seasons	Lilies	Roses	Tulips
Spring	182.08	237.55	183.37
Summer	178.76	233.22	180.02

Squared over Expected

Seasons	Lilies	Roses	Tulips
Spring	0.08	0.13	0.01
Summer	0.26	0.12	0.80

Performing a χ^2 Test of Independence

Compute the test statistic

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

Squared over Expected

Seasons	Lilies	Roses	Tulips
Spring	0.08	0.13	0.01
Summer	0.26	0.12	0.80
Autumn	0.02	0.02	0.08
Winter	0.01	0.33	0.55

Sum of Squared over Expected - χ^2

0.08 + 0.26 + 0.02 + ... + 0.08 + 0.55 =

```
## [1] 2.397417
```

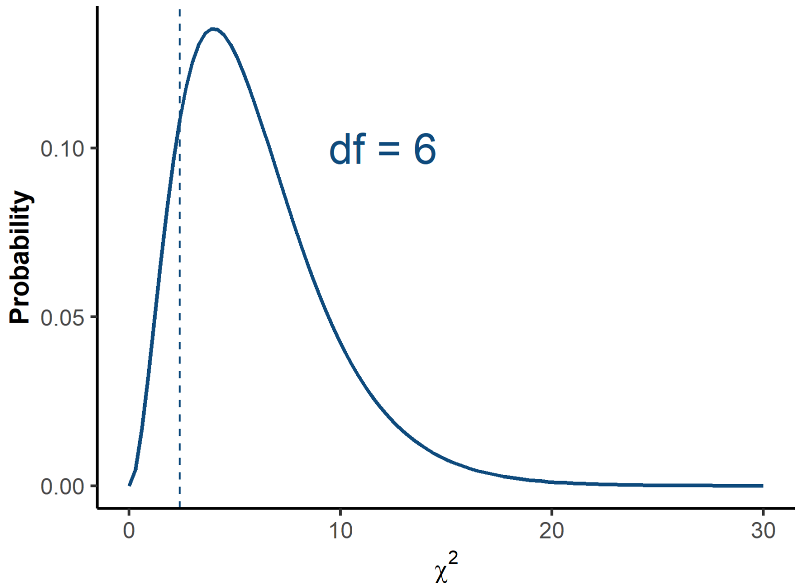
Performing a χ^2 Test of Independence

Find the test statistic on the distribution

$df = (r-1)(c-1)$

- where:
  - c = number of levels within Variable 1
  - r = number of levels within Variable 2

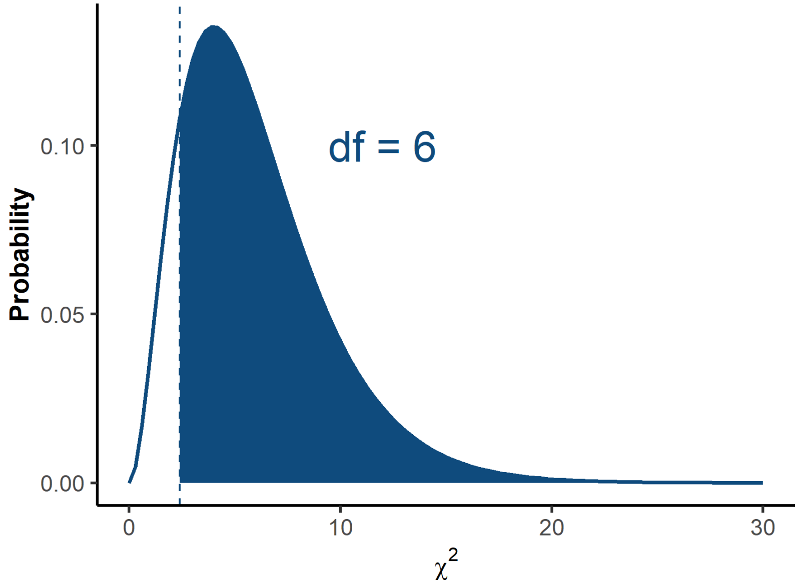
- so, in our example:
  - c = number of levels within Season
  - r = number of levels within Flowers
  - $df \;=\; (4-1)(3-1) \;=\; (3)(2) \;=\; 6$



Performing a χ^2 Test of Independence

Compute the probability of obtaining a χ^2 statistic at least as extreme as the observed one, if H_0 is true

- What proportion of the plot falls in the shaded area?



Performing a χ^2 Test of Independence

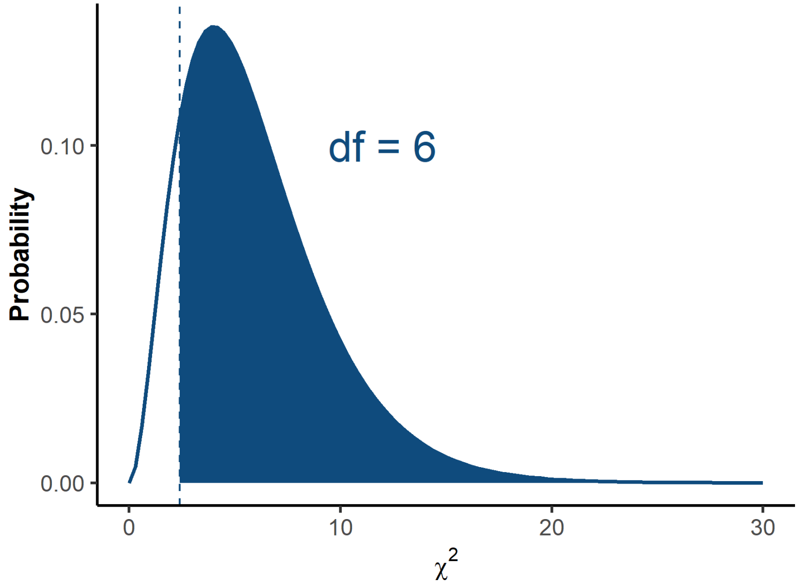
Compute the probability of obtaining a χ^2 statistic at least as extreme as the observed one, if H_0 is true

- What proportion of the plot falls in the shaded area?

```
pchisq(x2_stat_toi,
       df = 6,
       lower.tail = F)
```

```
## [1] 0.8797671
```

- The probability that we would have a χ^2 value as extreme as 2.4 if H_0 was true is 0.88



Performing a χ^2 Test of Independence in R

```
TOItest <- chisq.test(seasonDat$Season, seasonDat$Flowers)
TOItest
```

```
##
##      Pearson's Chi-squared test
##
## data:  seasonDat$Season and seasonDat$Flowers
## X-squared = 2.3974, df = 6, p-value = 0.8798
```

Exploring our Results Further

- We can compute standardized residuals for the Test of Independence
  - $\frac{O_{ij}-E_{ij}}{\sqrt{E_{ij}}}$

In R

```
TOItest$residuals
```

```
##                   seasonDat$Flowers
## seasonDat$Season      Lilies       Roses      Tulips
##           Spring  0.29040508 -0.36030121  0.12071137
##           Summer -0.50559019 -0.34179680  0.89285276
##           Autumn  0.12563386  0.13115144 -0.27447085
##           Winter  0.09053276  0.57407330 -0.74363023
```

By Hand

- You need to calculate separately by cell
- *Example of number of Lilies sold in Spring*
  - $\frac{O_{ij}-E_{ij}}{\sqrt{E_{ij}}} = \frac{186-182.08}{\sqrt{182.08}} = 0.291$

Write Up

A χ^2 test of independence was performed to examine whether the type of flower sold was independent of season. There was no significant association between these variables $(\chi^2(6, n = 2348) = 2.40, p = .880)$. Therefore, using an $\alpha = .05$, we failed to reject the null hypothesis.

Questions?

Effect Size

Effect Sizes

- There are 3 possibilities:
  - Phi coefficient (ϕ)
  - Cramer's V (V)
  - Odds Ratios (OR)
- You will learn more about odds ratios in DAPR2, so we will focus on Phi and Cramer's V

Phi Coefficient

$$\phi = \sqrt{\frac{\chi^2}{n}}$$

- where
  - n = total number of observations
- Should only be used when you have a 2x2 contingency table (2 categorical variables with 2 levels each)
- Common 'cut-offs' for ϕ-scores:

Verbal label	Magnitude of ϕ
Small effect	0.1
Medium effect	0.3
Large effect	0.5

Phi Coefficient in R

```
library(effectsize)
phi(TOItest)
```

Cramer's V

$$V = \sqrt{\frac{\chi^2}{n \cdot df^*}}$$

- where
  - n = total number of observations
  - df^* = $min(r-1, c-1)$
- Can be used when you aren't working with a 2x2 contingency table

Interpretation

- Cramer's V is interpreted based on df^*
- Cramer's V must lie between 0 and 1
  - 0 = complete independence, 1 = complete dependence

df^*	small	medium	large
1	.10	.30	.50
2	.07	.21	.35
3	.06	.17	.29
4	.05	.15	.25
5	.04	.13	.22

Cramer's V

In R

```
library(effectsize)
cramers_v(cont_table)
```

```
## Cramer's V (adj.) |       95% CI
## ---------------------------------
## 0.00              | [0.00, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].
```

By hand

$$V \quad = \quad \sqrt{\frac{\chi^2}{n \cdot df^*}} \quad = \quad \sqrt{\frac{2.40}{2348 \cdot 2}} \quad = \quad 0.023$$

Summary

- Today we have covered:
  - The χ^2 distribution and how it compares to the t distribution
  - The assumptions of χ^2 tests
  - How the χ^2 Goodness of Fit test and the χ^2 Test of Independence are different
  - How to calculate both types of χ^2 values
  - Standardized residuals and how they relate to your χ^2 results
  - Measures of effect size you may use with χ^2 tests

This Week

Tasks

- Attend both lectures
- Attend your lab and work on the assessed report with your group (due by 12 noon on Friday 28th of March 2025)
- Complete the weekly quiz
    - Opened Monday at 9am
    - Closes Sunday at 5pm

Support

- **Office Hours**: for one-to-one support on course materials or assessments
(see LEARN > Course information > Course contacts)
- **Piazza**: help each other on this peer-to-peer discussion forum
- **Student Adviser**: for general support while you are at university
(find your student adviser on MyEd/Euclid)