MEDB 5501, Module 13

2024-11-12

Topics to be covered

- What you will learn
  - The two by two crosstabulaton
  - Test of equality of two proportions
  - Sample size calculations
  - R code
  - Your homework

Listing of titanic.yaml, 1

```
---
data_dictionary: titanic.txt

description: >
  The Titanic was a large cruise ship, the biggest of its kind in 1912. It was
  thought to be unsinkable, but when it set sail from England to America in
its
  maiden voyage, it struck an iceberg and sank, killing many of the passengers
  and crew. You can get fairly good data on the characteristics of passengers
  who died and compare them to those that survived. The data indicate a strong
  effect due to age and gender, representing a philosophy of "women and
children
  first" that held during the boarding of life boats.

additional_description:
  http://www.statsci.org/data/general/titanic.html
```

Listing of titanic.yaml, 2

```
download_url:
  - http://www.statsci.org/data/general/titanic.txt

format:
  tab-delimited

varnames:
  first row of data

missing_value_code:
  NA
```

Listing of titanic.yaml, 3

```
size:
  rows: 1313
  columns: 5

vars:
  Name:
    label: Passenger name

  PClass:
    label: Passenger class
    scale: ordinal
    values: 1st, 2nd, 3rd
```

Listing of titanic.yaml, 4

```
Age:
  unit: years
  scale: ratio
  range: positive real numbers
  missing: NA

Sex:
  scale: binary
  values: female, male
```

Listing of titanic.yaml, 5

```
  Survived:
    scale: binary
    values:
      1: yes
      0: no
---
```

The crosstabulation of two binary variables

		Variable2	
		0	1
Variable1	0	n00	n01
	1	n10	n11

Speaker notes

One of the most common tables you will see in Statistics is the 2 by 2 crosstabulation. This table shows the counts associated with the combination of the levels of the two binary variables. There are only four numbers in this table, but there are numerous statistics that you can use to summarize what’s going on in this table.

Example: Titanic data

- Crosstabulation

```
        survived
sex      yes  no
  female 308 154
  male   142 709
```

Speaker notes

This is an example of a crosstabulation.The number in the upper left corner, 308, represents the number of female passengers who survived (did not die). This includes Kate Winslet. The number in the lower right corner, 709, represents then number of male passengers who did not survive. This includes, sad to say, Leonardo diCaprio.

Multiple ways to display, 1

- Crosstabulation

```
        survived
sex      yes  no
  female 308 154
  male   142 709
```

- Swap rows

```
        survived
sex      yes  no
  male   142 709
  female 308 154
```

- Swap columns

```
        survived
sex       no yes
  female 154 308
  male   709 142
```

- Swab both

```
        survived
sex       no yes
  male   709 142
  female 154 308
```

Speaker notes

The cross tabulation changes when you swap the rows, the columns, or both.

Multiple ways to display, 2

- Transposed

```
        sex
survived female male
     yes    308  142
     no     154  709
```

- Transposed, swap columns

```
        sex
survived male female
     yes  142    308
     no   709    154
```

- Transposed, swap rows

```
        sex
survived female male
     no     154  709
     yes    308  142
```

- Transposed, swab both

```
        sex
survived male female
     no   709    154
     yes  142    308
```

Speaker notes

You can get four more tables by transposing the matrix. What was the rows becomes the columns and what was the columns becomes the rows.

Row and column percents

- Crosstabulation with row totals

```
        survived
sex       yes  no Sum
  female 308 154 462
  male   142 709 851
```

- Row percents

```
        survived
sex              yes        no
Sum
  female 0.6666667 0.3333333
1.0000000
  male   0.1668625 0.8331375
1.0000000
```

- Cross tabulation with column totals

```
        survived
sex       yes  no
  female 308 154
  male   142 709
  Sum    450 863
```

- Column percents

```
        survived
sex              yes        no
  female 0.6844444 0.1784473
  male   0.3155556 0.8215527
  Sum    1.0000000 1.0000000
```

Speaker notes

The column percents are computed by dividing by the column totals. They add up to 100% within each column. The row percents are computed by dividing by the row totals. They add up to 100% within each row. The cell percents are computed by dividing by the overall total. They only add up to 100% when you sum across both the rows and the columns.

Cell percents

- Cell totals

```
        survived
sex        yes   no  Sum
  female   308  154  462
  male     142  709  851
  Sum      450  863 1313
```

- Cell percents

```
        survived
sex              yes        no       Sum
  female 0.2345773 0.1172887 0.3518660
  male   0.1081493 0.5399848 0.6481340
  Sum    0.3427266 0.6572734 1.0000000
```

My recommendation

- Outcome variable is the columns
- Intervention/exposure variable is the rows
- Calculate row percentages

```
        survived
sex            yes        no
  female 0.6666667 0.3333333
  male   0.1668625 0.8331375
```

Speaker notes

I have found that nine times out of ten, you want row percentages with the exposure/intervention variable as the rows and the outcome variable as the columns. This doesn’t always work, but it is usually what I try first. It shows how much the chances of a good outcomes (or sometimes the chances of a bad outcome) change when you switch levels of the exposure/intervention.

The orientation is also important. You want the percentages that are most interesting as close as possible. This is the proximity principle. The values within a column are nested above/below the other. The values within a row are farther apart.

Always round, 1

```
  0.6666667
- 0.1668625
  ---------
  ?????????
```

Speaker notes

Often in a cross tabulation, you will do some mental math. You might, for example, want to subtract survival probabilities.

Always round, 1

```
  0.67
- 0.17
  ----
  ????
```

Speaker notes

Notice how much easier the subtraction becomes when you round to two significant figures.

Break #1

- What you have learned
  - The two by two crosstabulaton
- What’s coming next
  - Test of equality of two proportions

The binomial distribution

- You have n trials of an “experiment”
- Two outcomes, “success” or “failure”
- \(\pi\) is probability of success
- Each trial is independent

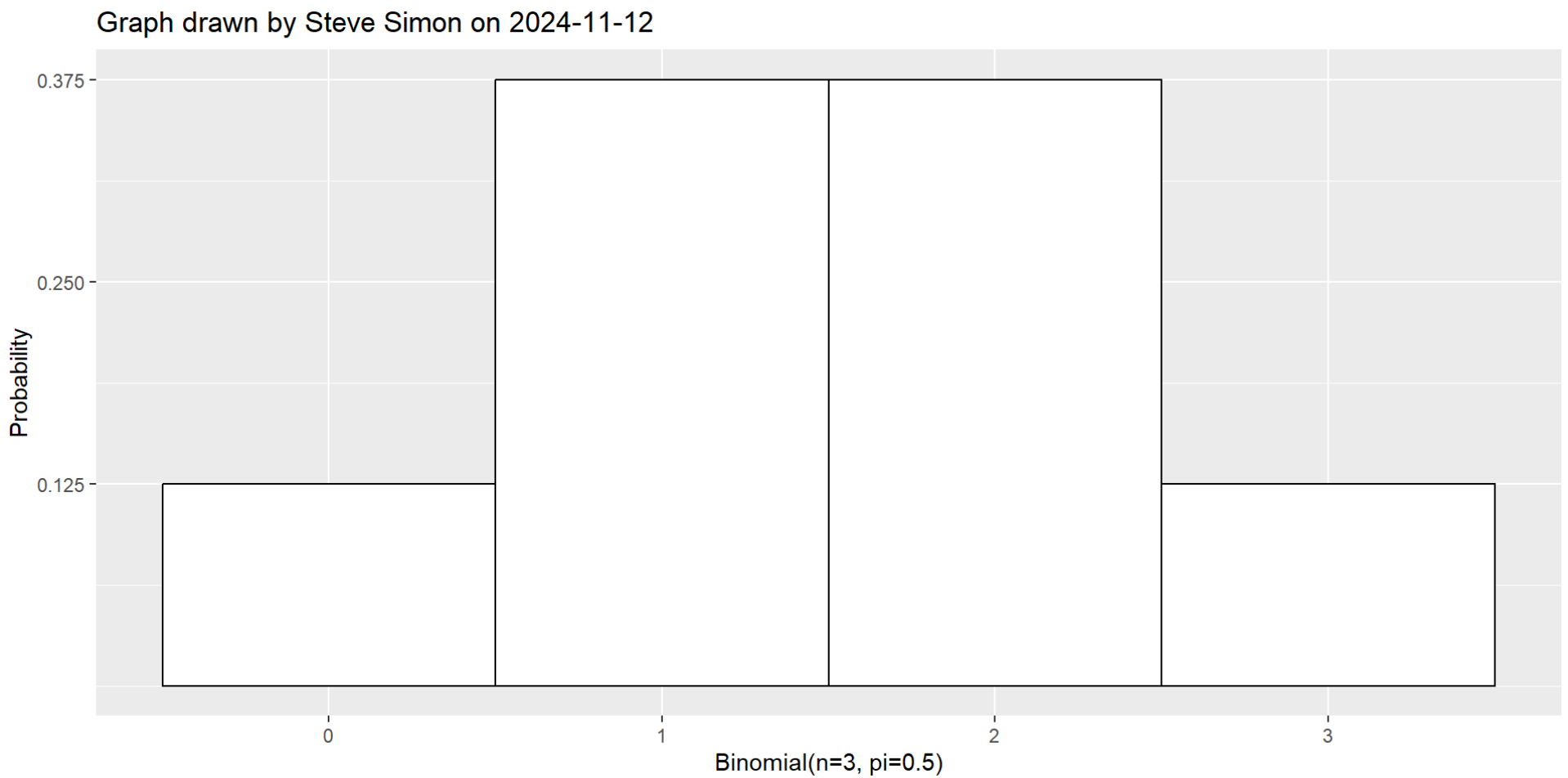
Example, creating a family of three children, 1

- You have 3 trials (pregnancies)
- Two outcomes, girl (success) or boy (failure)
- \(\pi\) = 0.5 is probability of success
- Each child is independent

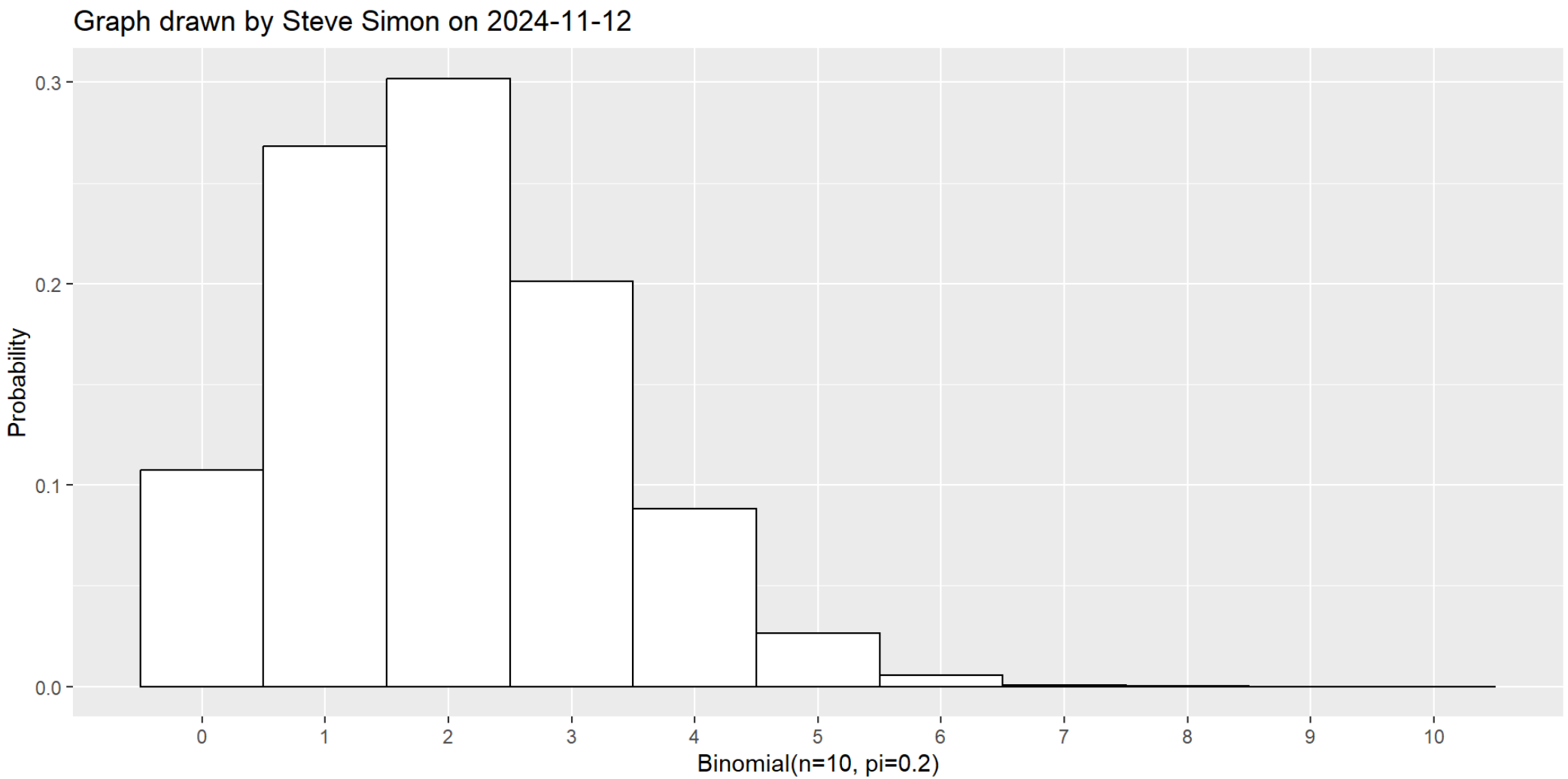
Example, creating a family of three children, 2

- Eight possible outcomes
  - X=0, BBB
  - X=1, BBG, BGB, GBB
  - X=2, BGG, GBG, GGB
  - X=3, GGG

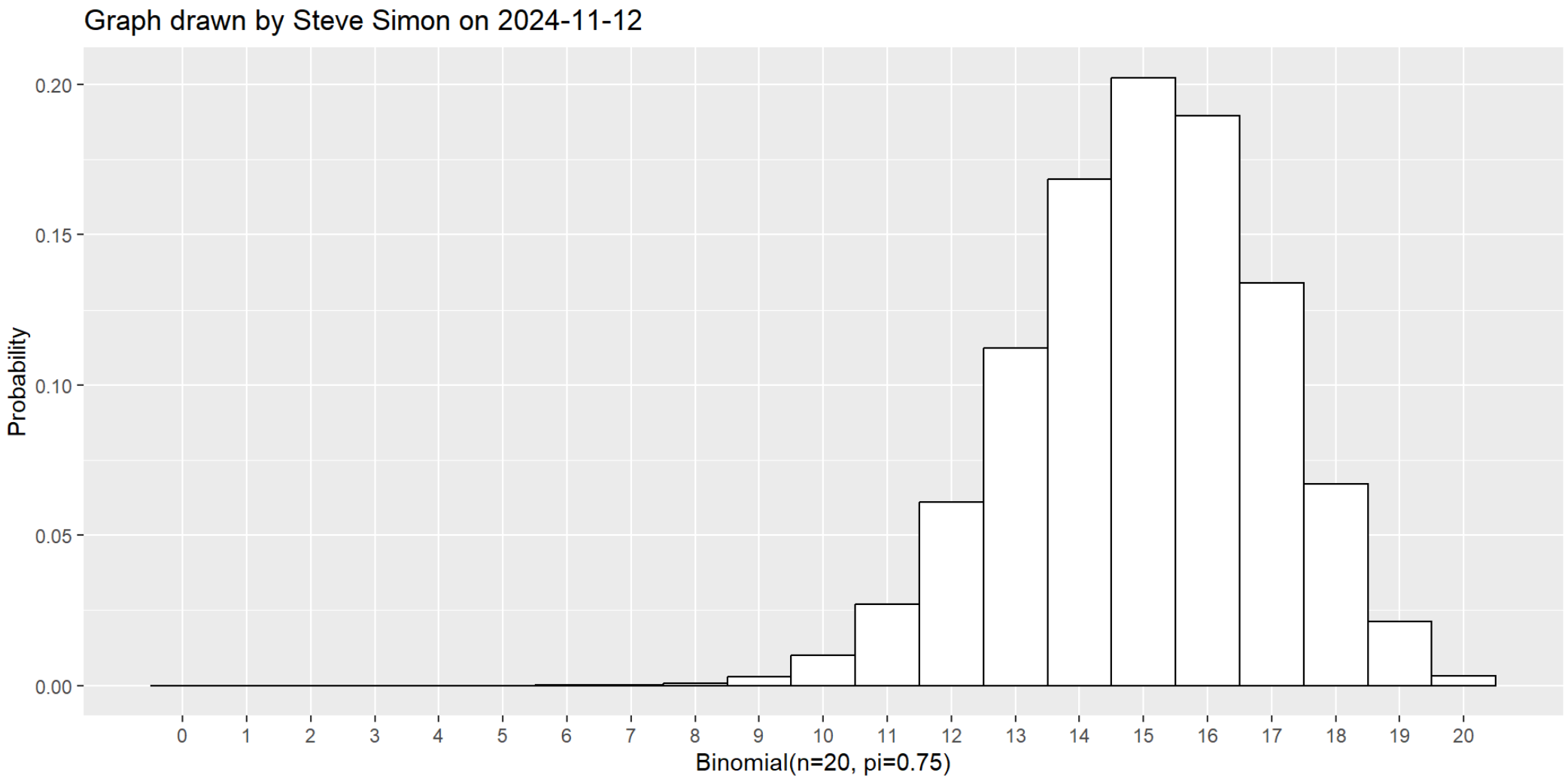
Example, creating a family of three children, 3



Example, Binomial(n=10, pi=0.2)



Example, Binomial(n=20, pi=0.75)



Two proportions model, 1

- Scenario 1
  - X_1 is a random binomial(n_1, π_1)
  - X_2 is a random binomial(n_2, π_2)
  - X_1 and X_2 are independent

Two proportions model, 2

- Scenario 2
  - Sample $X_{11}, X_{12}, ..., X_{1n_1}$
  - Sample $X_{21}, X_{22}, ..., X_{2n_2}$
  - Only possible values are 0, 1
  - $P[X_{1i}=1]=\pi_1$, $P[X_{2i}=1]=\pi_2$

Two proportions model, 3

- $H_0:\ \pi_1 - \pi_2 = 0$
- $H_1:\ \pi_1 - \pi_2 \ne 0$

Two proportions model, 4

- $T = \frac{p_1-p_2}{se}$
  - $se=\sqrt{\bar{p}(1-\bar{p})(\frac{1}{n_1}+\frac{1}{n_2})}$
  - $\bar{p}=\frac{n_1p_1+n_2p_2}{n_1+n_2}$
- Accept H_0 if T is close to zero.
  - If $Z(\alpha/2) < T < Z(1-\alpha/2)$

Two proportions model, 5

- p-value = \(2P[Z >|T|]\)
- Accept \(H_0\) if p-value > \(\alpha\)

Speaker notes

You can also compute a p-value which is the probability of observing the test statistic, T, or a value more extreme.

Some variations of the test statistic, 1

- Yates continuity correction
  - $T = \frac{p_1-p_2 + c}{se}$
  - Formula for c is messy
  - Net effect is to pull T closer to zero
  - Better to meet the normal approximation

Some variations of the test statistic, 2

- Chi-squared test
  - $T = \Big(\frac{p_1-p_2}{se}\Big)^2$
  - Accept H_0 if T < $\chi^2(1-\alpha, df=1)$
  - p-value = $P[\chi^2(df=1) > T]$
  - Does not allow for easy test of one-sided hypothesis
- Chi-squared test with Yates continuity correction
  - $T = \Big(\frac{p_1-p_2 + c}{se}\Big)^2$

One-sided test, 1

- $H_0:\ \pi_1 - \pi_2 = 0$
- $H_1:\ \pi_1 - \pi_2 > 0$
  - Accept H_0 if T < $z(1-\alpha)$
  - p-value = $P[Z > T]$
  - Accept H_0 if p-value > α

One-sided test, 2

- $H_0:\ \pi_1 - \pi_2 = 0$
- $H_1:\ \pi_1 - \pi_2 < 0$
  - Accept H_0 if T > $z(\alpha)$
  - p-value = $P[Z < T]$
  - Accept H_0 if p-value > α

Two sample test of proportions with Titanic data, 1

```
        survived
sex       yes  no
  female 308 154
  male   142 709
```

Two sample test of proportions with Titanic data, 2

```
        survived
sex            yes        no
  female 0.6666667 0.3333333
  male   0.1668625 0.8331375
```

Two sample test of proportions with Titanic data, 3

```
    2-sample test for equality of proportions without continuity correction

data:  table1
X-squared = 332.06, df = 1, p-value < 2.2e-16
alternative hypothesis: two.sided
95 percent confidence interval:
 0.4500519 0.5495564
sample estimates:
   prop 1    prop 2
0.6666667 0.1668625
```

Break #2

- What you have learned
  - Test of equality of two proportions
- What’s coming next
  - Sample size calculations

What you need for a continuous outcome

- Research hypothesis
- Standard deviation of your outcome measure
- Minimum clinically important difference
- Other details
  - Type I error rate (usually 0.05)
  - Power (usually 0.90)

Speaker notes

In an earlier lecture, you saw that a sample size calculation needed three major elements, a research hypothesis, a standard deviation, and the minimum clinically important difference.

That’s not all, but the Type I error rate or alpha is usually fixed at 0.05. Power is usually fixed at 0.9. You might go a bit lower, but don’t go much lower than 0.8 for power. It might be okay from your perspective, but power around 0.75 or less is considered a red flag by those who might examine your research (IRB members, granting agencies, journal peer reviewers).

What you need for a categorical outcome

- Research hypothesis
- Expected proportion in the control group
- Minimum clinically important difference

Speaker notes

The elements that you need when you have a categorical outcome change a bit. Instead of a standard deviation, you need to specify the expected proportion of events in the control group.

Hypothetical scenario, 1

- Weight loss study
  - Control is recommended diet and exercise routine
  - Treatment adds experimental weight loss drug
  - Binary event, losing at least 5 pounds after 6 months
- Hypothesis proportion is higher in treatment group than control
- Proportion in control group is 10%
- Need to see at least 5% more events in treatment group

Hypothetical scenario, 2

- Set power to 90%
- Set Type I error rate (alpha) to 5%

```
power.prop.test(
    p1=0.10,
    p2=0.15,
    power=0.9,
    sig.level=0.05)
```

```
     Two-sample comparison of proportions power calculation

              n = 917.3206
             p1 = 0.1
             p2 = 0.15
      sig.level = 0.05
          power = 0.9
    alternative = two.sided

NOTE: n is number in *each* group
```

Rule of 50

- Want to detect a doubling or halving
- Need 25 to 50 events in each group
- Example
  - Control probability is 0.1
  - Want to see 20% in treatment group
  - 25/0.1 = 250; 50/0.1 = 500

Break #3

- What you have learned
  - Sample size calculations
- What’s coming next
  - R code

Listing of simon-5501-13-titanic.qmd, 1

```
---
title: "Analysis of Titanic dataset"
format:
  html:
    embed-resources: true
---

This program reads data on survival of passengers on the Titanic. Find more
information in the [data dictionary][web00].

[web00]: https://github.com/pmean/data/blob/main/files/titanic.yaml

This code was written by Steve Simon on 2024-11-09 and is placed in the public
domain.
```

Listing of simon-5501-13-titanic.qmd, 2

````
## Load the tidyverse library

```{r}
#| label: setup
#| message: false
#| warning: false
library(broom)
library(epitools)
library(tidyverse)
```
````

Listing of simon-5501-13-titanic.qmd, 3

```
#### Comments on the code

For most of your programs, you should load the [tidyverse library][web01]. The
messages and warnings are suppressed.

[web01]: https://www.tidyverse.org/

In previous programs, I put a label for each chunk inside the curly braces
({}).
It is recommended instead to put the label on a separate line inside the
program
chunk. It is a bit more work to provide a unique label for each chunk, but it
helps quite a bit to isolate where to look when your code produces an error.
```

Listing of simon-5501-13-titanic.qmd, 4

````
## Read the data and view a brief summary

```{r}
#| label: read
ti <- read_tsv(
  file="../data/titanic.txt",
  col_names=TRUE,
  col_types="ccncn",
  na="NA")
names(ti) <- tolower(names(ti))
glimpse(ti)
```
````

Listing of simon-5501-13-titanic.qmd, 5

```
#### Comments on the code

Use read_tsv from the [readr package][web02] to  read this file. Use
col_names=TRUE because the column names are included as the first row of the
file. The col_types="ccncn" specifies the first second and fourth columns as
strings and the third and fifth as numeric. There are missing values in this
dataset, designated by the letters "NA".

[web02]: https://readr.tidyverse.org/
```

Listing of simon-5501-13-titanic.qmd, 6

````
## Replace numeric codes for survived

```{r}
#| label: replace-numbers
ti$survived <-
    factor(
        ti$survived,
        level=1:0,
        labels=c("yes", "no"))
```
````

Listing of simon-5501-13-titanic.qmd, 7

```
#### Comments on the code

The [factor function][web03] places the levels of a categorical variable in a
specific order and (optionally) attaches labels to each level. In this code,
the
number codes are reordered so that 1 appears first followed by 0. The labels
"yes" and "no" are attached to these two codes.

[web03]: https://stat.ethz.ch/R-manual/R-devel/library/base/html/factor.html

## Get counts of sex by survival

```{r}
#| label: counts
table1 <-xtabs(~sex+survived, data=ti)
```
```

Listing of simon-5501-13-titanic.qmd, 8

```
#### Comments on the code

The [table function][web04] or the [xtabs function][web05] creates a matrix
with the number of observations in each combination of sex and survived. The
[count function][web06] provides an alternative where these values are placed
in a single column.

[web04]: https://stat.ethz.ch/R-manual/R-devel/library/base/html/table.html
[web05]: https://stat.ethz.ch/R-manual/R-devel/library/stats/html/xtabs.html
[web06]: https://dplyr.tidyverse.org/reference/count.html

#### Interpretation of the output

There are 154 female passengers who died and 308 who survived. There are 709
male passengers who died and 142 who survived.
```

Listing of simon-5501-13-titanic.qmd, 9

````
## Get proportions for died/survived by sex

```{r}
#| label: proportions
table1 |>
  proportions("sex")
```


#### Comments on the code

The [proportions function][pro1] calculate proportions for a matrix, either by
row (margin=1), by column (margin=2) or across all cells (margin=NULL). This
code shows an alternative where you specify which variable to summarize
across. The line proportions("sex") tells R to compute totals for each level
of sex and divide by those totals. In this case, since sex represents the rows
````

Listing of simon-5501-13-titanic.qmd, 10

```
#### Interpretation of the output

The proportion of women who died is 33%. The proportion of men who died is
much higher at 83%.
```

Listing of simon-5501-13-titanic.qmd, 11

````
## Bar chart

```{r}
#| label: bar
ti |>
  ggplot() +
    aes(x = sex, fill = survived) +
    geom_bar(position = "fill") +
      xlab("Sex") +
      ylab("Proportion") +
      ggtitle("Graph drawn by Steve Simon on 2024-11-11")
```

#### Comments on the code
````

Listing of simon-5501-13-titanic.qmd, 12

````
#### Interpretation of the output

I am not a big fan of bar charts, but they are quite popular. This bar chart shows a large disparity in the proportion who survived in each sex (the upper portion of the bar in peach). Equivalently, you can say that there is a large disparity in the proportion who died in each sex (the lower portion of the bar in green-blue).

## Hypothesis test for two proportions

```{r}
#| label: equality-of-proportions
prop.test(table1, correct=FALSE)
```
````

Listing of simon-5501-13-titanic.qmd, 13

```
#### Comments on the code

The [prop.test function][pro2] does not work on raw data. You need to get
summary counts, either from the table function or the xtabs function. The
option correct=FALSE informs R to not use the [Yates continuity correction]
[yat1].

[pro2]: https://search.r-project.org/R/refmans/stats/html/prop.test.html
[yat1]: https://www.statskingdom.com/121proportion_normal2.html
```

Listing of simon-5501-13-titanic.qmd, 14

```
#### Interpretation of the output.

The Chi-squared statistic is much larger than the degrees of freedom and the
p-value is small. So you would reject the null hypothesis and conclude that
there is a statistically significant difference in the mortality rates between
men and women on the Titanic. The 95% confidence interval for the difference
in proportions is -0.55 to -0.45. This interval excludes the value of zero and
indicates that the mortality rate is at least 45% lower and possibly as much
as
55% lower for men.
```

Listing of simon-5501-13-titanic.qmd, 15

````
## Power calculation

```{r}
#| label: power
power.prop.test(p1=0.17, p2=0.67, power=0.9)
```
````

Listing of simon-5501-13-titanic.qmd, 16

```
#### Comments on the code

The [power.prop.test][pow1] function will compute a sample size if you specify
the two group probabilities, the significance level, and the power. It will
also
do other calculations, such as the minimum difference in probabilities that
you
could detect for a specific sample size and a specific power.

[pow1]: https://stat.ethz.ch/R-manual/R-
devel/library/stats/html/power.prop.test.html
```

Listing of simon-5501-13-titanic.qmd, 17

```
#### Interpretation of the output

Clearly, a sample size of 1,313 passengers provided more than enough power and
precision, especially considering that the difference in survival rates was so
large. Out of curiosity, what sample size would still be adequate for testing
a
shift of 50% from a survival rate of 17% to a survival rate of 67%? If you
wanted power to be at least 90%, then a sample size of only 19 passengers of
each sex would be needed.
```

Listing of simon-5501-13-titanic.qmd, 18

````
## Chi-squared test

```{r}
#| label: chi-square-test
chisq.test(table1, correct=FALSE)
```
````

Listing of simon-5501-13-titanic.qmd, 19

```
#### Comments on the code

The [chisq.test function][chi1] calculates a chi-square test of independence.
It
takes input in a variety of forms. In this example, it uses a crosstabulation
computed by the xtabs command as input.

This function also will run a goodness-of-fit test, which is not discussed in
this lecture.

[chi1]: https://stat.ethz.ch/R-manual/R-
devel/library/stats/html/chisq.test.html
```

Listing of simon-5501-13-titanic.qmd, 20

````
#### Interpretation of the output

The chi-squared statistic is much larger than the degrees of freedom and the
p-value is small. You should reject the null hypothesis and conclude that sex
and survival are related (not independent)

## Fisher's Exact test

```{r}
#| label: fishers-exact
fisher.test(table1)
```
````

Listing of simon-5501-13-titanic.qmd, 21

```
#### Comments on the code

The [fisher.test function][fis1] calculates the Fisher's exact test, which is
helpful for small sample sizes. The 1,313 passengers on the Titanic do not
constitute a small sample size by any means. This test is just shown as an
example of how to calculate this test.

[fis1]: https://stat.ethz.ch/R-manual/R-
devel/library/stats/html/fisher.test.html
```

Listing of simon-5501-13-titanic.qmd, 22

````
#### Interpretation of the output

The p-value is small. You should reject the null hypothesis and conclude that
sex and survival are related (not independent). The estimated odds ratio is
9.97. The confidence interval for the odds ratio excludes the value of 1
leading to the same conclusion. In fact, even after allowing for sampling
error
that odds of survival are at least 7.6 times greater for women than for me.

## Odds ratio calculation

```{r}
#| label: odds-ratio
oddsratio(table1)
```
````

Listing of simon-5501-13-titanic.qmd, 23

```
#### Comments on the code

The oddsratio function and riskratio function (see below) are part of the
[epitools library][epi1]. It produces an odds ratio and confidence interval
and p-values associated with the Fisher's Exact test and the Chi-squared test
of
independence.

[epi1]: https://cran.r-project.org/web/packages/epitools/epitools.pdf
```

Listing of simon-5501-13-titanic.qmd, 24

```
#### Interpretation of the output

We are 95% confident that the odds ratio of survival for women versus men is
at
least 7.7 and possibly as large as 13, after accounting for sampling error.
This interval excludes the value of 1, so you can conclude that the risk of
death is much higher for men than for women. Equivalently you could conclude
that the odds of survival are much higher for women than for men.
```

Listing of simon-5501-13-titanic.qmd, 25

````
## Risk ratio calculation, 1

```{r}
#| label: risk-ratio-1
table1 |>
  proportions("sex")
```

#### Interpretation of the output

Before calculating the risk ratio, let's look at the row percentages one more time. The probability of survival is around 2/3 for women and about 1/6 for men. This means that the risk ratio from a survival perspective is around 4 (2/3 divided by 1/6). The probability of death is 1/3 for females and about 5/6 for males. The risk ratio from a mortality perspective is 0.4 (1/3 divided by 5/6)
````

Listing of simon-5501-13-titanic.qmd, 26

````
## Risk ratio calculation, 2

```{r}
#| label: risk-ratio-2
riskratio(table1)
```

#### Interpretation of the output

The risk ratio is comparing the probability of death between men and women. Men have 2.5 times higher probability of death compared to women. The confidence interval excludes the value of 1, indicating a statistically significant increase.
````

Listing of simon-5501-13-titanic.qmd, 27

````
## Risk ratio calculation, 3

```{r}
#| label: risk-ratio-3
riskratio(table1, rev="columns")
```

#### Interpretation of the output

The risk ratio is comparing the probability of survival between men and women. Men has one-fourth the probability of survival compared to women. The confidence interval excludes the value of 1, indicating that men have a statistically significantly lower probability of survival compared to women.
````

Listing of simon-5501-13-titanic.qmd, 28

````
## Save data for later use

```{r save}
save(ti, file="../data/titanic.RData")
```

#### Comments on the code

It is usually a good idea to [save][sav1] your data in an RData file to make
it
easier to retrieve this data later (with the [load function][loa1]).

[sav1]: https://stat.ethz.ch/R-manual/R-devel/library/base/html/save.html
[loa1]: https://stat.ethz.ch/R-manual/R-devel/library/base/html/save.html
````

Break #4

- What you have learned
  - R code
- What’s coming next
  - Your homework

Listing of simon-5501-13-directions.md, 1

```
---
title: "Directions for 5501-13 programming assignment"
---

This programming assignment was written by Steve Simon on 2024-10-08 and is
placed in the public domain.
```

Listing of simon-5501-13-directions.md, 2

```
## Program

-   Download the [program][tem]
    -   Store it in your src folder
-   Modify the file name
    -   Use your last name instead of "simon"
-   Modify the documentation header
    -   Add your name to the author field
    -   Optional: change the copyright statement

[tem]: https://github.com/pmean/classes/blob/master/biostats-1/13/src/simon-5501-13-titanic.qmd
```

Listing of simon-5501-13-directions.md, 3

```
## Data

-   Download the [data][dat] file
    -   Store it in your data folder
-   Refer to the [data dictionary][dic], if needed.

[dat]: https://github.com/pmean/data/blob/main/files/titanic.txt
[dic]: https://github.com/pmean/data/blob/main/files/titanic.yaml
```

Listing of simon-5501-13-directions.md, 4

````
## Question 1

Create a new variable, third_class that indicates whether a passenger is in
third class or not. The code would look something like this.

```{}
ti$third_class <-
  case_when(
    ti$pclass == "1st" ~ "no",
    ti$pclass == "2nd" ~ "no",
    ti$pclass == "3rd" ~ "yes")
```

How many passengers were in third class?
````

Listing of simon-5501-13-directions.md, 5

```
## Question 2

What are the probabilities of survival for third class passengers. How does
this
compare to the probability of survival for the other passengers.

## Question 3

Test the hypothesis that the survival probability is different for third class
passengers and the other passengers. Interpret the p-value and confidence
interval.
```

Listing of simon-5501-13-directions.md, 6

```
## Your submission

-   Save the output in html format
-   Convert it to pdf format.
-   Make sure that the pdf file includes
    -   Your last name
    -   The number of this course
    -   The number of this module
-   Upload the file

## If it doesn't work

Please review the [suggestions if you encounter an error page][sim3].

[sim3]: https://github.com/pmean/classes/blob/master/general/suggestions-if-
```

Summary

- What you have learned
  - The two by two crosstabulaton
  - Test of equality of two proportions
  - Sample size calculations
  - R code
  - Your homework