

# $t$ -Distribution Examples

CSU, Chico Math 314

2018-10-12

# outline

Bernoulli

Normal

References

# outline

Bernoulli

Normal

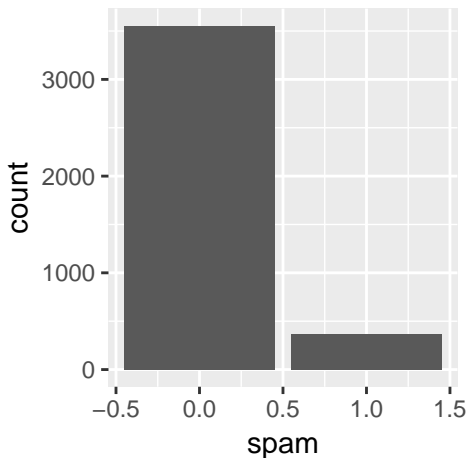
References

## Proportion of emails spam

Consider the `email` data set. Calculate and interpret a 98% confidence interval for the proportion of emails that are spam.

## Proportion of emails spam

```
email <- read.csv("https://roualdes.us/data/email.csv")  
ggplot(data=email, aes(spam)) + geom_bar()
```



## Proportion of emails spam

```
phat <- mean(email$spam)
s <- sd(email$spam)
n <- length(email$spam)
tstar <- qt(c(0.01, 0.99), df=n-1)

phat + tstar*s/sqrt(n)

## [1] 0.08277167 0.10442547
```

## Proportion of emails spam, conclusion

We are 98% confident that the true proportion of emails that are spam is between 0.08 and 0.1.

## Proportion of emails spam, extra

Follow up questions.

1. What sort of issues with these data should we keep in mind?
2. Does the interval above include the true population proportion of interest?
3. Would a 90% confidence interval be wider or narrower? 99%?
4. How could you use this interval to conclude the following hypothesis test?

$$H_0 : p = 0.15 \text{ versus } H_1 : p \neq 0.15; \alpha = 0.02$$



# outline

Bernoulli

Normal

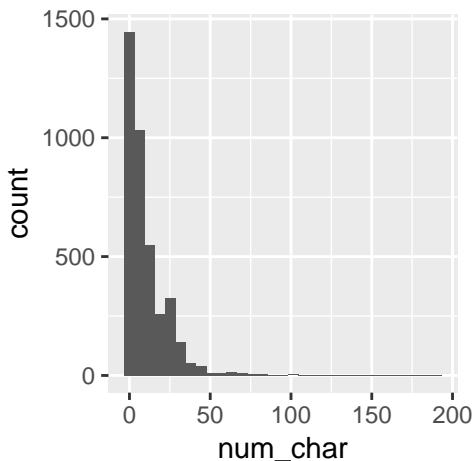
References

## Mean number of characters per email

Consider the `email` data set. Calculate and interpret a 95% confidence interval for the mean number of characters per email.

## Mean number of characters per email

```
ggplot(data=email, aes(num_char)) +  
  geom_histogram(bins=31)
```



## Mean number of characters per email

```
xbar <- mean(email$num_char)
s <- sd(email$num_char)
n <- length(email$num_char)
tstar <- qt(c(0.025, 0.975), df=n-1)

xbar + tstar*s/sqrt(n)

## [1] 10.24803 11.16515
```

## Mean number of characters per email, extra

We are 95% confident that the true mean number of line characters per email is between 10.25 and 11.17.

## Mean number of characters per email, extra

Follow up questions.

1. What sort of issues with these data should we keep in mind?
2. Does the interval above include the true population mean of interest?
3. Does more confidence mean a wider interval or smaller?
4. How could you use this interval to conclude the following hypothesis test?

$$H_0 : \mu = 11 \text{ versus } H_1 : \mu \neq 11; \alpha = 0.05$$

# outline

Bernoulli

Normal

References

## references I

- David M Diez, Christopher D Barr, and Mine Cetinkaya-Rundel.  
*OpenIntro Statistics*. CreateSpace independent publishing platform, third edition, 2015.
- Hadley Wickham. *ggplot2: elegant graphics for data analysis*. Springer Science and Business Media, 2009.