ENV 710: Lecture 5

confidence intervals & p-values

Duke | NICHOLAS SCHOOL OF THE ENVIRONMENT

roadmap

- where we are?
- lab assignments/reports
- questions?
- pod work!

descriptive statistics

discrete probability/distributions

continuous probability/distributions

↓

inference

sampling, central limit theorem, confidence intervals, t-distribution, p-values

↓

one- and two-sample tests

z-test, t-tests, etc., more on hypothesis testing and statistical power

tips for lab write-ups

1. Clearly label your answers. Read through your final knitted pdf as if you were the grader -- make sure your code is organized and its clear what part of the question you are addressing.
2. Read through the question and make sure you answer ALL parts.
3. No need to write anything beyond what is asked. If the question asks for the probability of x, just state "The probability of x is…".
4. Not necessary to include a written explanation of your coding unless explicitly stated in the question (Graders can read your code and annotations in your appendix).
5. Go to office hours! We are here to help.

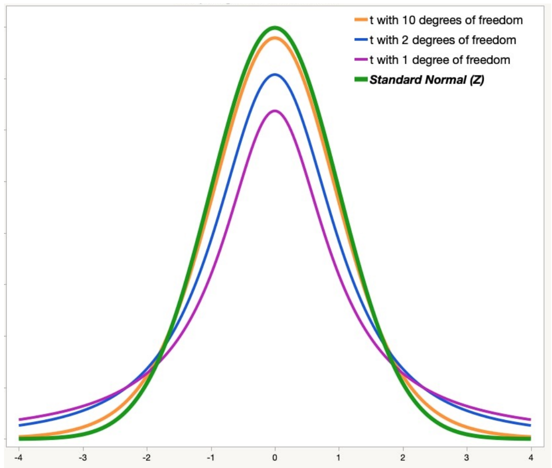
summary

confidence intervals

$$\bar{x} \pm Z_{\alpha/2} \frac{s}{\sqrt{n}}$$

$$\bar{x} \pm t_{[n-1]\alpha/2}(\frac{s}{\sqrt{n}})$$

t distribution

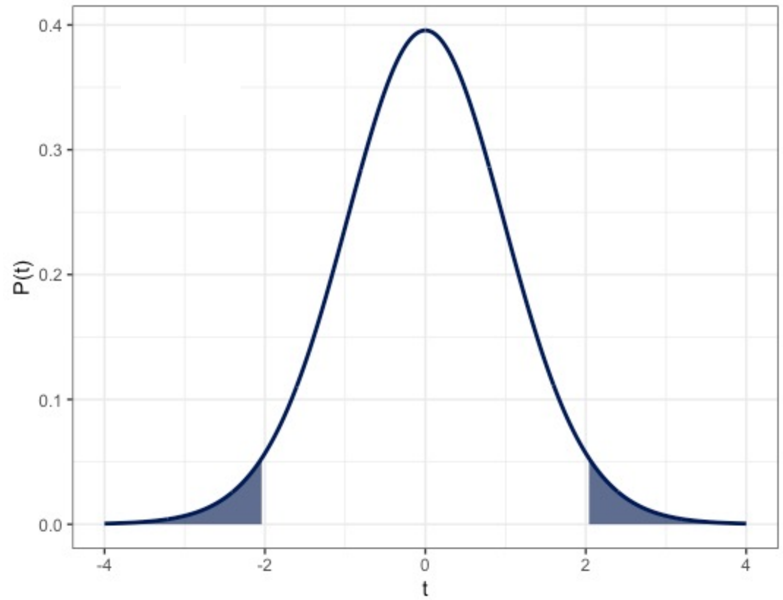


p-values

P(observed or more extreme outcome|H_0 true)

QUESTIONS & ANSWERS

```
abs(qt(0.025, df = 9-1))
qt(0.975, df = 9-1)
[1] 2.306004
```



Z or *t* are statistics, that can be used to obtain probabilities

- given our H_0, what is the probability of getting data as extreme or more extreme than our data (represented by a statistic)

1 – summary

Discuss the following

- what is a confidence interval? what does it mean to say that the average temperature in July 2050 will be 110 °F with a 95% CI of 107 to 113°F?
- what is a *p*-value? what does it tell us? with an alpha (significance level) of 0.05, what would we conclude about our null hypothesis if p = 0.03 or p = 0.24?

Based on a random sample of 100 vehicles, a 90% CI for the mean speed of vehicles on Circuit Drive is calculated to be (29.5 mph, 32.5 mph). Which is true?

(a) 90% of all vehicles on Circuit Dr. drive at speeds between 29.5 and 32.5 mph.

(b) we are 90% confident that the interval (29.5 mph, 32.5 mph) captures the true mean speed of all vehicles on Circuit Dr..

(c) we are 90% confident that a randomly selected vehicle will have a speed between 29.5 and 32.5 mph.

(d) the mean speed of the vehicles is 31.0 mph 90% of the time.

(e) 90% of all samples will have mean speeds between 29.5 and 32.5 mph.

I – summary

Discuss the following

- what is a confidence interval? what does it mean to say that the average temperature in July 2050 will be 110 °F with a 95% CI of 107 to 113°F?
- what is a *p*-value? what does it tell us? with an alpha (significance level) of 0.05, what would we conclude about our null hypothesis if $p = 0.03$ or $p = 0.24$?

Based on a random sample of 100 vehicles, a 90% CI for the mean speed of vehicles on Circuit Drive is calculated to be (29.5 mph, 32.5 mph). Which is true?

(a) 90% of all vehicles on Circuit Dr. drive at speeds between 29.5 and 32.5 mph.

(b) we are 90% confident that the interval (29.5 mph, 32.5 mph) captures the true mean speed of all vehicles on Circuit Dr..

(c) we are 90% confident that a randomly selected vehicle will have a speed between 29.5 and 32.5 mph.

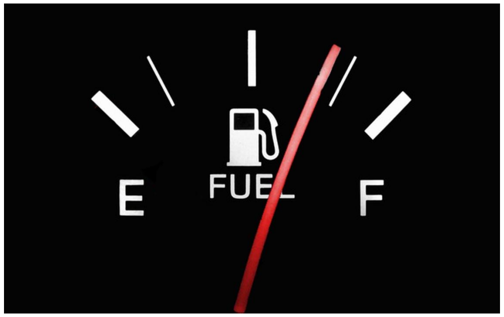
(d) the mean speed of the vehicles is 31.0 mph 90% of the time.

(e) 90% of all samples will have mean speeds between 29.5 and 32.5 mph.

(10 min.)

2 – car models

A random sample of 75 car models was taken to evaluate highway gas mileage. The cars in the sample had an average gas mileage of 26.1 mpg, with a standard deviation of 6.07. The sample distribution was only slightly right skewed.



1. calculate a 95% confidence interval around the mean gas mileage
2. articulate what the 95% confidence interval tells you
3. calculate a 90% confidence interval around the mean gas mileage
4. what sample size and/or distribution conditions must be met before you can calculate the confidence interval?
5. calculate a 95% confidence interval if we only had a sample of 15 car models
6. compare/contrast the three different CI's

2 – car models

A random sample of 75 car models was taken to evaluate highway gas mileage. The cars in the sample had an average gas mileage of 26.1 mpg, with a standard deviation of 6.07. The sample distribution was only slightly right skewed.

$N = 75$

$\bar{x} = 26.1$

$s = 6.07$

$$\bar{x} \pm z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

$$SE = \frac{s}{\sqrt{n}} = \frac{6.07}{\sqrt{75}} = 0.70$$

$$26.1 \pm 1.96 \cdot 0.70 = \quad 26.1 \pm 1.372$$

$$(24.73, 27.47)$$

1. calculate a 95% confidence interval around the mean gas mileage

```
26.1 - qnorm(p = 0.975, mean = 0, sd = 1)*6.07/sqrt(75)
26.1 + qnorm(p = 0.975, mean = 0, sd = 1)*6.07/sqrt(75)
```

```
24.73, 27.47
```

2. articulate what this confidence interval tells you. – we are confident that the true (population) mpg lies within our confidence interval 90 or 95% of the time... or we are 90/95% confident

2 – car models

A random sample of 75 car models was taken to evaluate highway gas mileage. The cars in the sample had an average gas mileage of 26.1 mpg, with a standard deviation of 6.07. The sample distribution was only slightly right skewed.

3. calculate a 90% confidence interval around the mean gas mileage

```
26.1 - qnorm(p = 0.95, mean = 0, sd = 1)*6.07/sqrt(75)
26.1 + qnorm(p = 0.95, mean = 0, sd = 1)*6.07/sqrt(75)
```

```
24.95, 27.25
```

4. what sample size and/or distribution conditions must be met before you can calculate the confidence interval?

   random sample & 75 models is < 10% of all car models - *we assume the car models are independent of each other*

   N ≥ 30, and sample is only slightly skewed - *we assume the sampling distribution is nearly normal*

2 – car models

A random sample of 75 car models was taken to evaluate highway gas mileage. The cars in the sample had an average gas mileage of 26.1 mpg, with a standard deviation of 6.07. The sample distribution was only slightly right skewed.

5. calculate a 95% confidence interval if we only had a sample of 15 car models

```
26.1 - qt(p=0.975, df=15-1)*(6.07/sqrt(15))
26.1 + qt(p=0.975, df=15-1)*(6.07/sqrt(15))
```

```
22.73855, 29.46145
```

6. compare/contrast the three different CI's

95% CI: 24.73, 27.47

90% CI: 24.95, 27.25

95% $CI_{t\text{-}dist}$: 22.76, 29.44

90% CI is narrower than 95% CI
small sample 95% CI is wider than both

3 – car models

Based on a 95% CI of [24.7, 27.5], do the data support the hypothesis that cars on average have a higher gas mileage than 25.5 mpg?

Test the hypothesis, that given our data, cars on average have a higher gas mileage than 25.5 mpg (i.e., find the probability of the null hypothesis given our data).

$$N = 75$$

$$\bar{x} = 26.1$$

$$s = 6.07$$

(10 min.)

3 – car models

Based on a 95% CI of [24.7, 27.5], do the data support the hypothesis that cars on average have a higher gas mileage than 25.5 mpg?

Test the hypothesis, that given our data, cars on average have a higher gas mileage than 25.5 mpg (i.e., find the probability of the null hypothesis given our data).

$$N = 75$$

$$\bar{x} = 26.1$$

$$s = 6.07$$

(10 min.)

1 Question asks "what is the probability of getting our data, $\bar{x}$ = 26.1, given that μ = 25.5"

2 State the hypotheses

$$H_0 : \bar{x} = 25.5$$

$$H_a : \bar{x} > 25.5$$

$$P(\bar{x} > 25.5 \mid H_0 : \mu = 25.5)$$

3 – car models

Test the hypothesis, that given our data, cars on average have a higher gas mileage than 25.5 mpg (i.e., find the probability of the null hypothesis given our data).

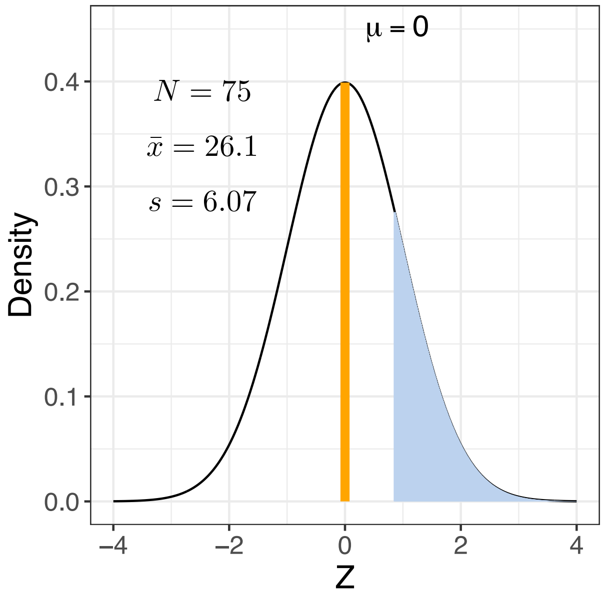
$$SE = \frac{s}{\sqrt{n}} = \frac{6.07}{\sqrt{75}} = 0.70$$

$$Z = \frac{26.1 - 25.5}{6.07/\sqrt{75}} = 0.856$$

$$P(Z > 0.856) = 0.196$$

```
z <- (26.1-25.5)/(6.07/sqrt(75))
1-pnorm(z, mean = 0, sd = 1)
[1] 0.1959883
```

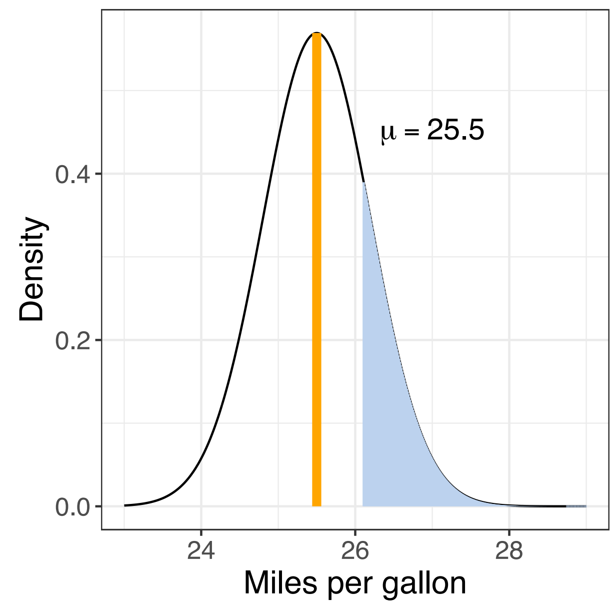
$$P(\bar{x} > 25.5 \mid H_0 : \mu = 25.5)$$



3 – car models

Test the hypothesis, that given our data, cars on average have a higher gas mileage than 25.5 mpg (i.e., find the probability of the null hypothesis given our data).

$$P(\bar{x} > 25.5 \mid H_0 : \mu = 25.5)$$



Or, wrapping it all into one R expression

```
1-pnorm(q=26.1, mean=25.5, sd=6.07/sqrt(75))
[1] 0.1959883
```

what does this mean in terms of our H_0?
do we reject or retain our H_0?

Questions?