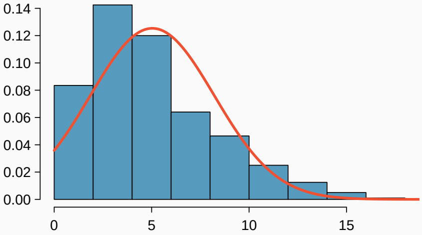
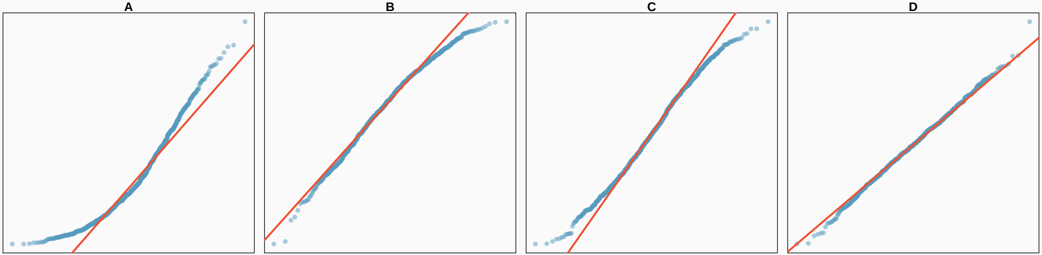
Announcements

- Homework 5 due on Monday
- I apologize, I wasn't able to get through all the test grading. : (Expect them Monday.
- In-class lab next Monday on Confidence Intervals
- Start reading Ch 3 for next Wednesday
- Polling: `rembold-class.ddns.net`

Review Question

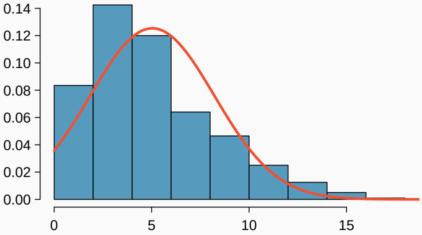
Which of the QQ plots below would best describe the data distribution to the right?

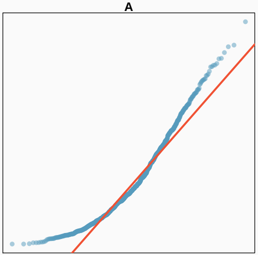




Review Question

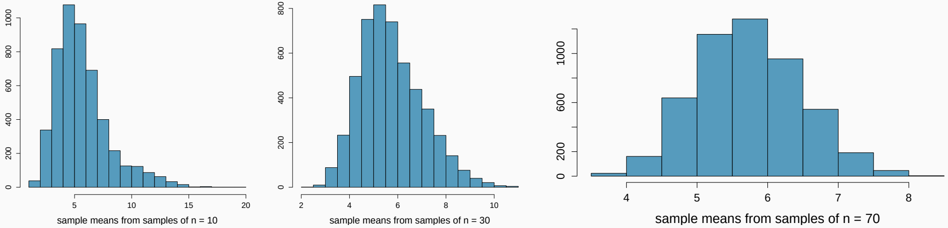
Which of the QQ plots below would best describe the data distribution to the right?





Back to our oddity

- When thinking about this standard error, some sample size dependence must factor in!
  - The deviations got smaller with larger sample sizes



- We'd like to pin this down now

Standard Error for Means

- That deviations get smaller with larger samples sizes suggests some sort of $\frac{1}{n}$ dependence
- End up using a $\frac{1}{\sqrt{n}}$ dependence owning to the square rooting of the variance
- Thus we have that

$$SE_{means} = \frac{s}{\sqrt{n}} = \frac{\text{sample standard dev}}{\sqrt{\text{sample number}}}$$

Example

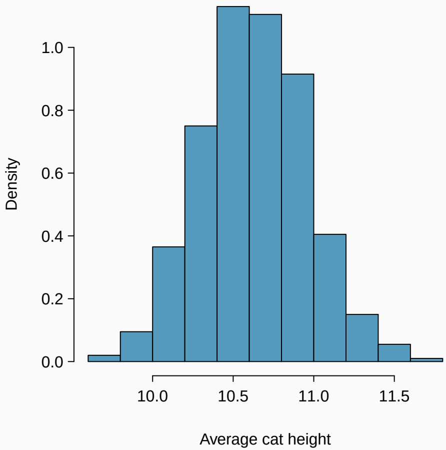
Suppose I'd just like to look at my data for the heights of cats from a particular data set. Suppose I take a random sample of 40 cats. And find that, in my sample:

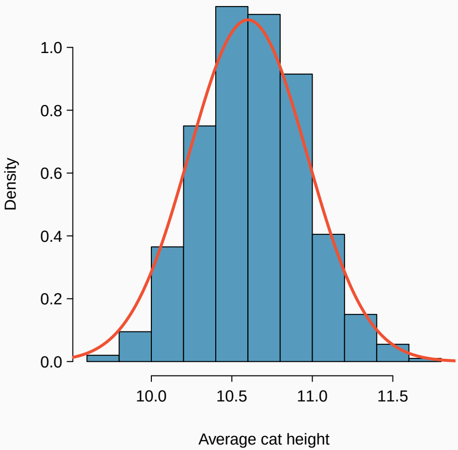
$\bar{x}$	s
10.60	2.32

What is my standard error of my sample mean?

Crafting the Normal Distribution

We use the standard error as our deviation to craft a normal distribution which we hope will approximate that of the true cat height distribution.





An absolute unit

Suppose you discover a new breed of cat which you measure to have an average height of 11.3 cm. Not being sure about average cat heights, you took a random sample of 40 cats of all manner of breeds and got the same data as on the previous slide. You are curious if the heights of this new cat breed are normal. What would you conclude?

An absolute unit

Suppose you discover a new breed of cat which you measure to have an average height of 11.3 cm. Not being sure about average cat heights, you took a random sample of 40 cats of all manner of breeds and got the same data as on the previous slide. You are curious if the heights of this new cat breed are normal. What would you conclude?

- Want two tail test
- One tail would give value of 2.8%
- Two tail doubles to put us at 5.6% which is greater than the normal 5% statistical significance value.
- Fail to reject null hypothesis

With Confidence

- Our point estimate is our best guess for the value of the population parameter
- There will almost always be some error in this value though
- We would like to instead describe some range of value in which we can be reasonable sure the population parameter lives
- Such a range is called a ***confidence interval***

More Tradeoffs

- A huge interval
  - You are extremely sure your population parameter is within the interval
  - But it is kinda useless
- A tiny interval
  - You are less confident that your population parameter is within that range
  - But it is very useful
- Confidence intervals (and the degree of confidence you have) will always be working within these tradeoffs

The 95% Confidence Interval

- A point estimate we observe will be within 1.96 standard errors of the true population parameter about 95% of the time
- A ***95% confidence interval*** is constructed then as

$$\text{point estimate} \pm 1.96 \times \text{SE}$$

- We interpret this as saying: “We are 95% sure this interval captures the true population parameter”

Example

Let's construct the 95% confidence interval for our cat height sample from earlier.

$\bar{x}$	s	SE
10.60	2.32	0.367

Example

Let's construct the 95% confidence interval for our cat height sample from earlier.

$\bar{x}$	s	SE
10.60	2.32	0.367

The mean population cat height is 10.63. Does it fall within our interval?

Understanding Check

Suppose you sample 20 of your friends and find that your average GPA is a 3.2 with a standard deviation of 0.3. What interval below best describes a 95% confidence interval in which the campus-wide average GPA score should fall?

A) 2.61 - 3.788

B) 2.915 - 3.485

C) 3.068 - 3.33

D) 3.17 - 3.2294

Understanding Check

Suppose you sample 20 of your friends and find that your average GPA is a 3.2 with a standard deviation of 0.3. What interval below best describes a 95% confidence interval in which the campus-wide average GPA score should fall?

A) 2.61 - 3.788
B) 2.915 - 3.485
C) 3.068 - 3.33
D) 3.17 - 3.2294