Announcements

- Homework 5 posted and due on Monday
- Still hoping to have test results to you on Friday
- Read Ch 2.8 on Confidence Intervals for Friday
- In-class lab next Monday on Confidence Intervals
- Polling: `rembold-class.ddns.net`

Review Question!

Means and standard deviations for some common normal distributions are given in the table to the right. Barry is a male with a height of 63 inches, while Susan is a female with a height of 68 inches. Jill and Jane wear women's size 6 and 11 shoes, respectively. Which of the 4 is statistically the most abnormal?

	μ	σ
Male height (inches)	70	4
Female height (inches)	65	3.5
Women's shoe sizes	8	1.5

A) Barry

B) Susan

C) Jill

D) Jane

Review Question!

Means and standard deviations for some common normal distributions are given in the table to the right. Barry is a male with a height of 63 inches, while Susan is a female with a height of 68 inches. Jill and Jane wear women's size 6 and 11 shoes, respectively. Which of the 4 is statistically the most abnormal?

	μ	σ
Male height (inches)	70	4
Female height (inches)	65	3.5
Women's shoe sizes	8	1.5

A) Barry

B) Susan

C) Jill

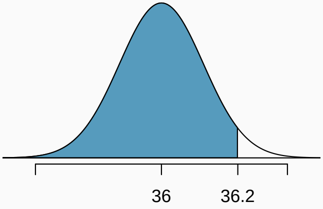
D) Jane

Back to Heinz

At Heinz ketchup factory the amounts which go into bottles of ketchup are supposed to be normally distributed with mean 36 oz. and standard deviation 0.11 oz. Once every 30 minutes a bottle is selected from the production line, and its contents are noted precisely. If the amount of ketchup in the bottle is below 35.8 oz. or above 36.2 oz., then the bottle fails the quality control inspection. What percentage of bottles pass inspection?

Back to Heinz

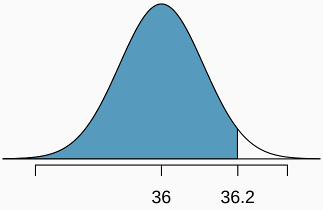
At Heinz ketchup factory the amounts which go into bottles of ketchup are supposed to be normally distributed with mean 36 oz. and standard deviation 0.11 oz. Once every 30 minutes a bottle is selected from the production line, and its contents are noted precisely. If the amount of ketchup in the bottle is below 35.8 oz. or above 36.2 oz., then the bottle fails the quality control inspection. What percentage of bottles pass inspection?



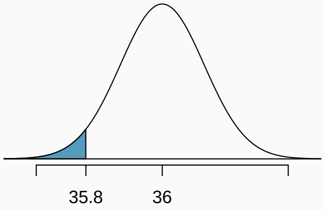
-

Back to Heinz

At Heinz ketchup factory the amounts which go into bottles of ketchup are supposed to be normally distributed with mean 36 oz. and standard deviation 0.11 oz. Once every 30 minutes a bottle is selected from the production line, and its contents are noted precisely. If the amount of ketchup in the bottle is below 35.8 oz. or above 36.2 oz., then the bottle fails the quality control inspection. What percentage of bottles pass inspection?



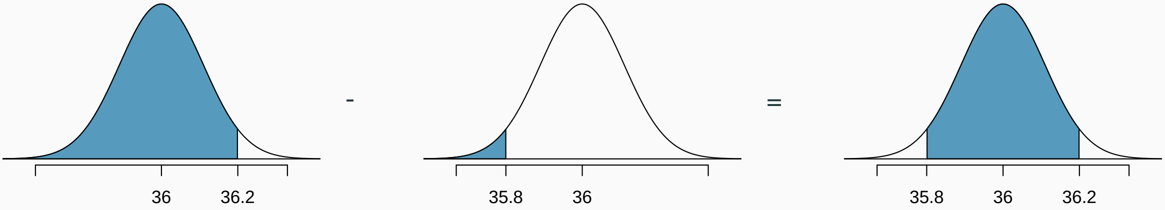
-



=

Back to Heinz

At Heinz ketchup factory the amounts which go into bottles of ketchup are supposed to be normally distributed with mean 36 oz. and standard deviation 0.11 oz. Once every 30 minutes a bottle is selected from the production line, and its contents are noted precisely. If the amount of ketchup in the bottle is below 35.8 oz. or above 36.2 oz., then the bottle fails the quality control inspection. What percentage of bottles pass inspection?



Finding the exact probability - using R

```
> pnorm(36.2, mean = 36, sd = 0.11)
[1] 0.9654818
```

OR

```
> pnorm(1.82, mean = 0, sd = 1)
[1] 0.9654818
```

Finding the exact probability - using R

```
> pnorm(36.2, mean = 36, sd = 0.11)
[1] 0.9654818
```

OR

```
> pnorm(1.82, mean = 0, sd = 1)
[1] 0.9654818
```

```
> pnorm(35.8, mean = 36, sd = 0.11)
[1] 0.0345
```

OR

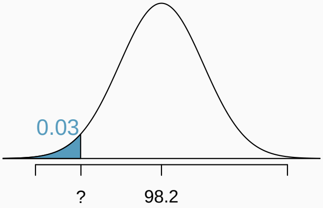
```
> pnorm(-1.82, mean = 0, sd = 1)
[1] 0.0344
```

Going the other direction

Body temperatures of healthy humans are distributed nearly normally with mean 98.2 ℉ and standard deviation 0.73 ℉. What is the cutoff for the lowest 3% of human body temperatures?

Going the other direction

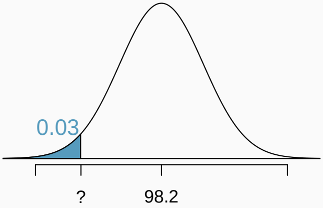
Body temperatures of healthy humans are distributed nearly normally with mean 98.2 °F and standard deviation 0.73 °F. What is the cutoff for the lowest 3% of human body temperatures?



0.09	0.08	0.07	0.06	0.05	Z
0.0233	0.0239	0.0244	0.0250	0.0256	−1.9
0.0294	0.0301	0.0307	0.0314	0.0322	−1.8
0.0367	0.0375	0.0384	0.0392	0.0401	−1.7

Going the other direction

Body temperatures of healthy humans are distributed nearly normally with mean 98.2 °F and standard deviation 0.73 °F. What is the cutoff for the lowest 3% of human body temperatures?



0.09	0.08	0.07	0.06	0.05	Z
0.0233	0.0239	0.0244	0.0250	0.0256	−1.9
0.0294	0.0301	0.0307	0.0314	0.0322	−1.8
0.0367	0.0375	0.0384	0.0392	0.0401	−1.7

```
> qnorm(0.03, mean = 98.2, sd = 0.73)
[1] 96.82702
```

Practice

Body temperatures of healthy humans are distributed nearly normally with mean 98.2 °F and standard deviation 0.73 °F. What is the cutoff for the highest 10% of human body temperatures?

A) 97.3 °F

B) 99.1 °F

C) 99.4 °F

D) 99.6 °F

Z	0.05	0.06	0.07	0.08	0.09
1.0	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9115	0.9131	0.9147	0.9162	0.9177

Practice

Body temperatures of healthy humans are distributed nearly normally with mean 98.2 ℉ and standard deviation 0.73 ℉. What is the cutoff for the highest 10% of human body temperatures?

A) 97.3 ℉

B) 99.1 ℉

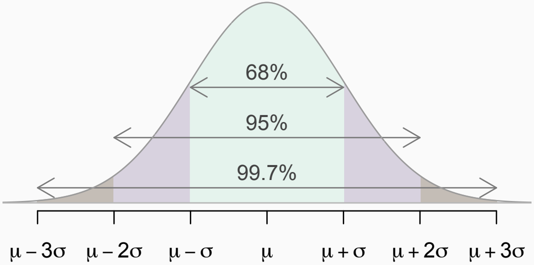
C) 99.4 ℉

D) 99.6 ℉

Z	0.05	0.06	0.07	0.08	0.09
1.0	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9115	0.9131	0.9147	0.9162	0.9177

68-95-99.7 Rule

- For nearly normally distributed data,
  - about 68% falls within 1 SD of the mean,
  - about 95% falls within 2 SD of the mean,
  - about 99.7% falls within 3 SD of the mean.
- It is possible for observations to fall 4, 5, or more standard deviations away from the mean, but these occurrences are very rare if the data are nearly normal.



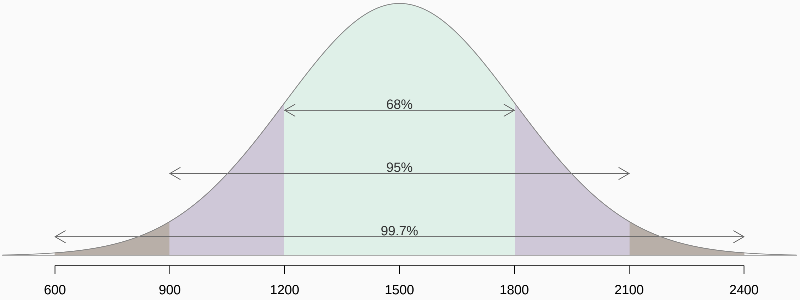
Describing variability using the 68-95-99.7 Rule

SAT scores are distributed nearly normally with mean 1500 and standard deviation 300.

Describing variability using the 68-95-99.7 Rule

SAT scores are distributed nearly normally with mean 1500 and standard deviation 300.

- ~68% of students score between 1200 and 1800 on the SAT.
- ~95% of students score between 900 and 2100 on the SAT.
- ~99.7% of students score between 600 and 2400 on the SAT.



Checking Normality

- We've stated that many things follow approximately normal distributions
- How can we actually check if something is following a normal distribution though?
    - Take our residuals from linear regression. We wanted to ensure they were largely normal. Besides a visual comparison, how can we check this?

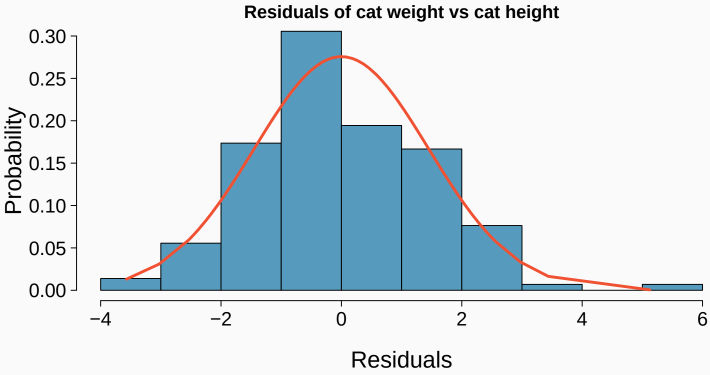
Checking Normality

- We've stated that many things follow approximately normal distributions
- How can we actually check if something is following a normal distribution though?
    - Take our residuals from linear regression. We wanted to ensure they were largely normal. Besides a visual comparison, how can we check this?
- Methods:
    - Visual check
    - Checking 68/95/99.7% rule
    - Normal probability plots

Visual Check of Normality

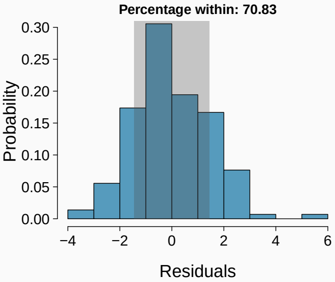
Overlay a normal plot over a histogram

- Use the sample mean $\bar{x}$ and standard deviation s to parameters for best fitting normal curve



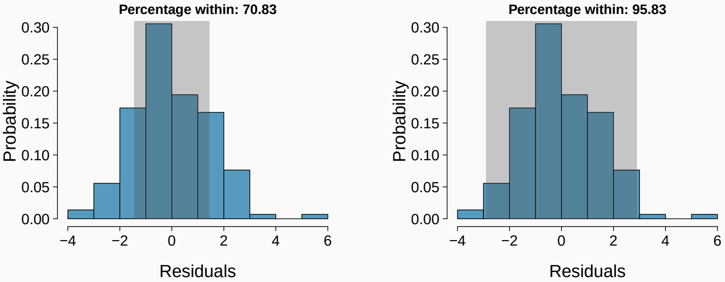
Checking 68-95-99.7% Rule

Can also directly compute the amount of observations within each standard deviation of the mean



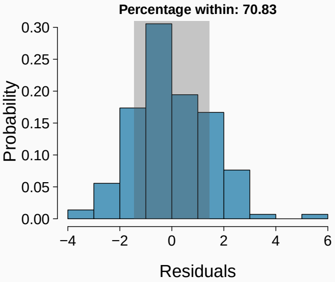
Checking 68-95-99.7% Rule

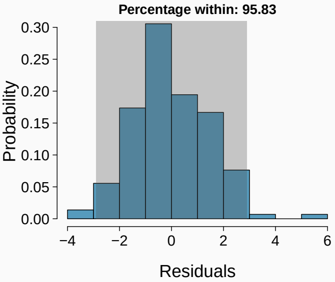
Can also directly compute the amount of observations within each standard deviation of the mean

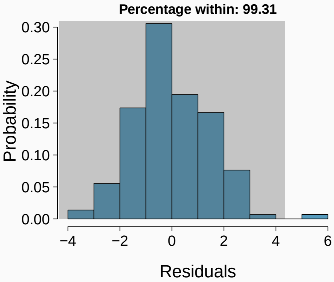


Checking 68-95-99.7% Rule

Can also directly compute the amount of observations within each standard deviation of the mean

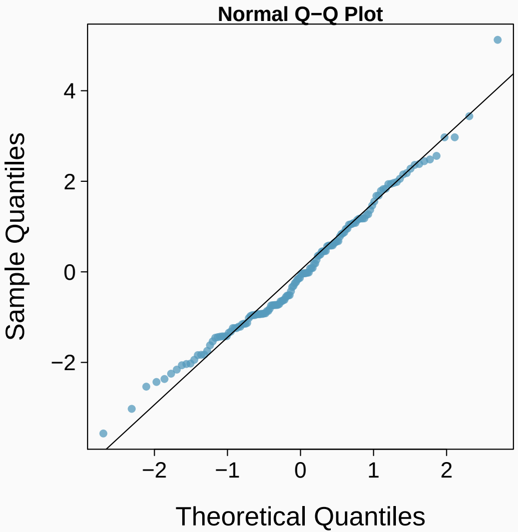






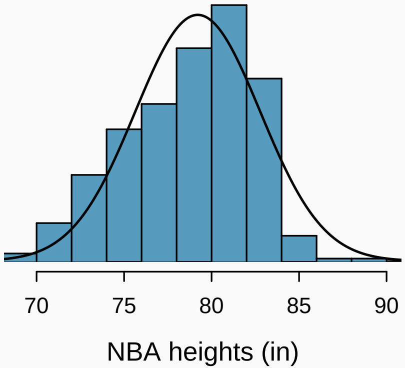
Checking a Normal Probability Plot

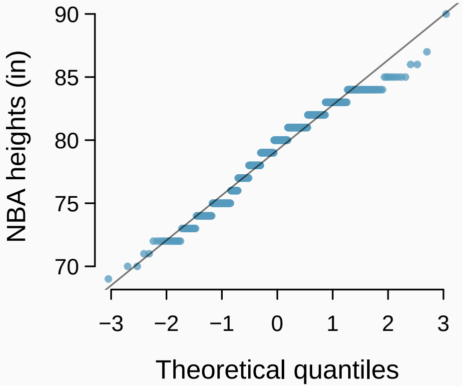
- Also called a ***quantile-quantile*** plot or ***qq*** plot
- Data on y-axis and theoretical normal quantiles on x-axis
- The closer everything lies on a line, the better it approximates a normal distribution
- Tedious to construct without a computer



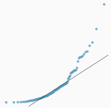
NBA Heights

Below is a histogram and normal probability plot for the NBA heights for the 2008-2009 season. Do these data appear to follow a normal distribution?

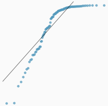




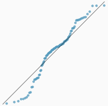
Q-Q Plots and Skewness



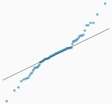
Right skew - Points bend up and to the left of the line



Left skew - Points bend down and to the right of the line



Short tails (narrower distribution) - Points follow an S-shaped curve



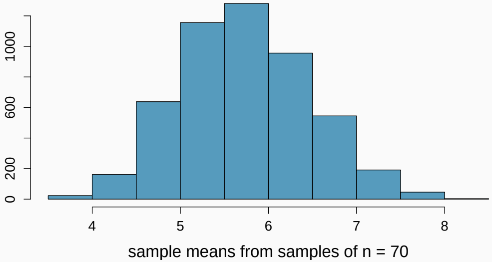
Long tails (broader distribution) - Points start below line, bend to follow, and then end above

Applications to Hypothesis Testing

- Originally we created null distributions through simulation
- Now can create through a normal distribution
- So long as the sample size is sufficiently large, we'll arrive at the same conclusions

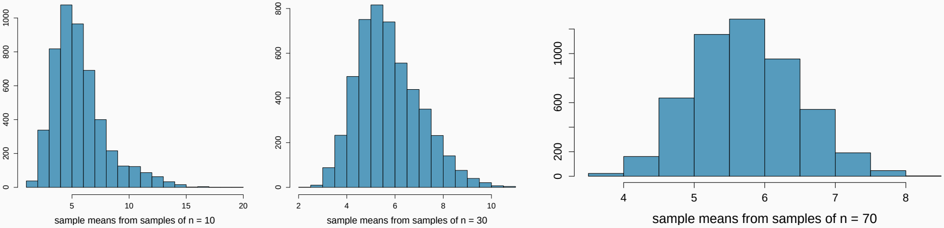
A question of error

- Our estimates for the mean of some population had some spread as seen in our sample distributions
- The variability of this estimate we call the ***standard error*** (SE).
- We need to know (or estimate) the standard error in order to be able to convert our simulation to a normal distribution
  - What is the correct standard deviation to use?



Something weird going on…

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



- We'll focus more on this in Ch 3