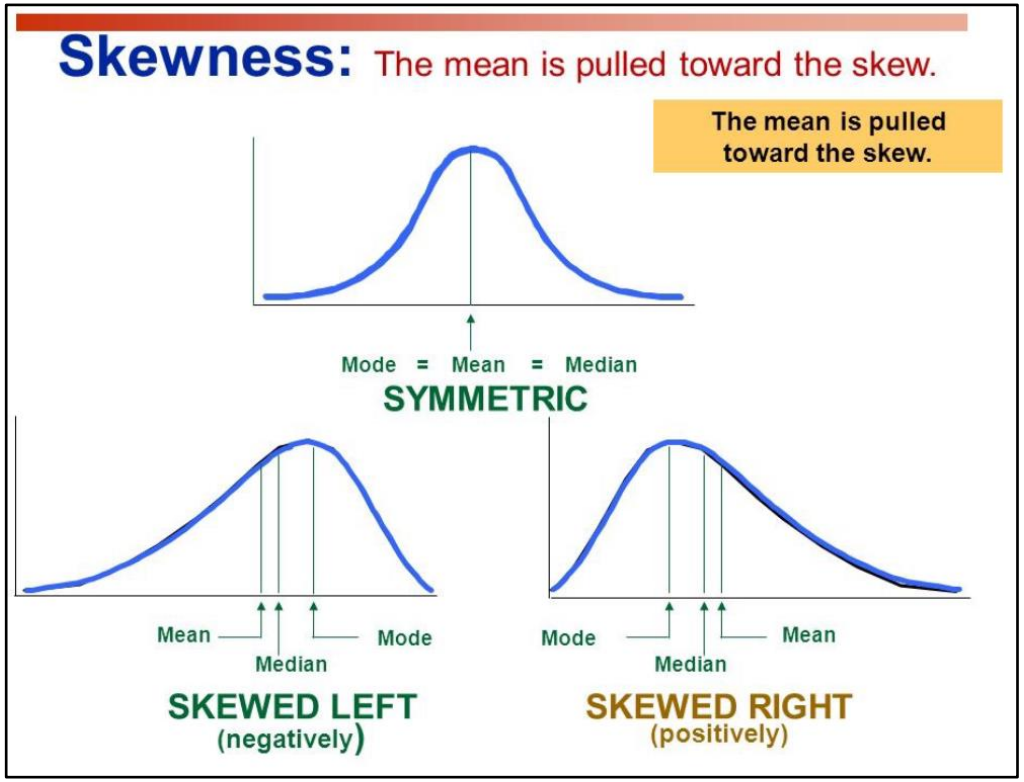
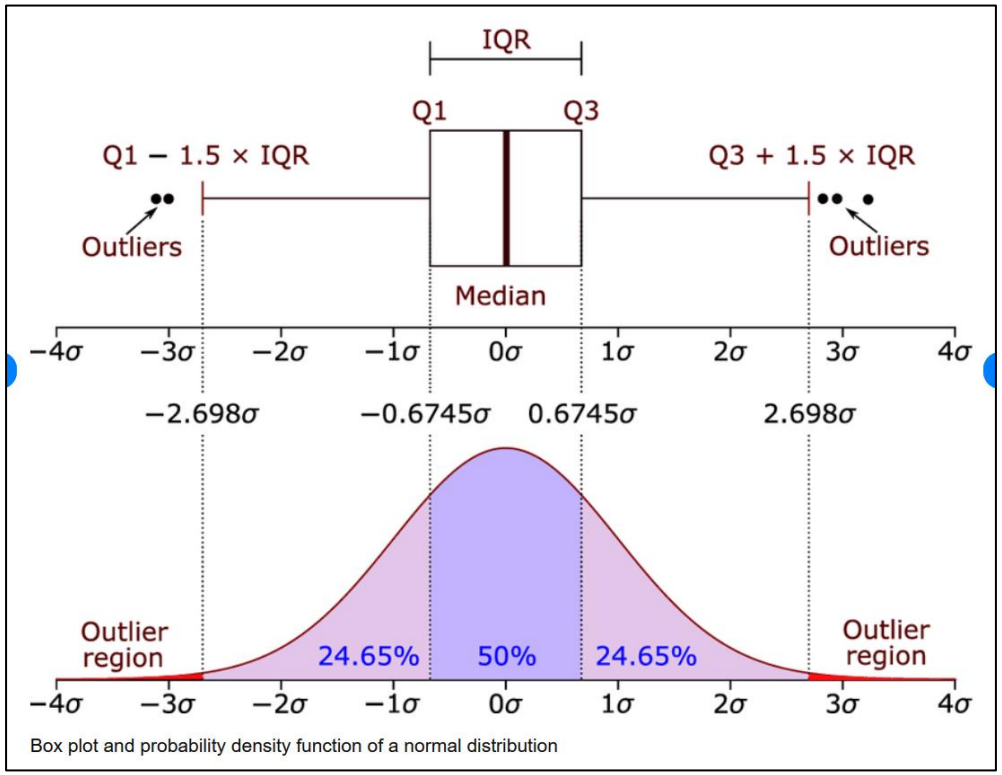
Central Tendency



- Mean is affected by outliers
- Median is not affected by outliers
- One way to tell skew is comparing mean to median
  - See how mean goes towards tail?
  - Outliers are in the tails.

Box Plots + SD



Box plot and probability density function of a normal distribution

- Q2 is the median and the split point of data set
  - If median is closer to Q3 or Q1 it is skewed to that side.
- IQR is where most of the data lives
- IQR = Q3 – Q1
- Whiskers are calculated to be thresholds for outlier data
- Standard deviations is a single calculated value for a single unit of spread. SD is in the unit of the data set.

Expected Value + Return Volatility

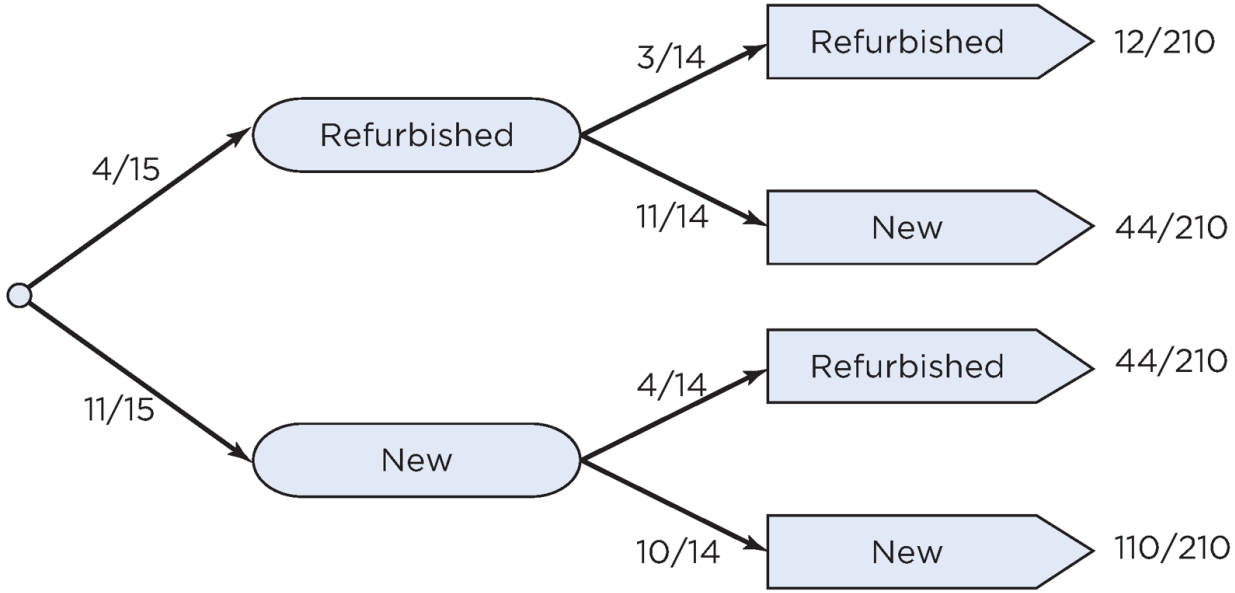
CheapO Profits. CheapO Computers shipped two servers to its biggest client. Four refurbished servers were mistakenly restocked along with 11 new systems.

If the client receives two new servers, the profit for the company is $10,000; if the client receives one new server, the profit is $4,500. If the client receives two refurbished systems, the company loses $1000.

What are the expected value and standard deviation of CheapO's profits?

Expected Value + Return Volatility

CheapO Profits. CheapO Computers shipped two servers to its biggest client. Four refurbished computers were mistakenly restocked along with 11 new systems.



	Revenue	Step 1	Step 2	Prob of final event	Final Prob	(Rev * Prob)	((Rev - EV)^2) * Prob
Both Refurbished	$ (1,000.00)	4/15 Chance	3/14 Chance	4/15*3/14	0.057	$ (57.14)	$ 3,718,349.21
One Refurbished	$ 4,500.00	4/15 Chance	11/14 Chance	(4/15*11/14)*2 (*2 because there are 2 occurances where this can happen)	0.419	$ 1,885.71	$ 2,760,592.59
None Refurbished	$10,000.00	11/15 Chance	10/14 Chance	11/15*10/14	0.524	$ 5,238.10	$ 4,507,089.95
					Expected Value (sum of rev*prob)	$ 7,066.67	
					Return Volatility (sqrt(sum of ((Rev - EV)^2) * Prob))	3314.518328	

Z-Scores and Probability

$$Z = \frac{x - \mu}{\sigma}.$$

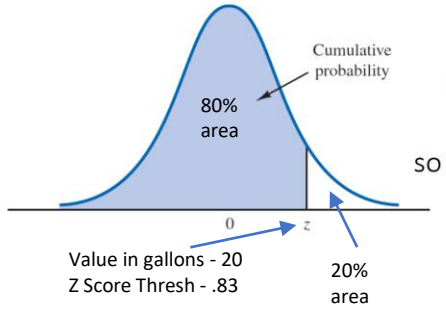
What is the probability of a stockout at Pep Zone?

- Pep Zone sells a popular multi-grade motor oil.
- When the stock of this oil drops to 20 gallons, a replenishment is ordered.
- The store manager is concerned that sales are being lost due to stockouts while waiting for a replenishment order.
- Demand during this time is normally distributed with $\mu = 15$ gallons and $\sigma = 6$ gallons.

$$P(X > 20) = ?$$

$$P(X > 20) = P\left(Z > \frac{20 - 15}{6}\right) = P(Z > 0.83)$$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389



We know $P(Z \le 0.83) = 0.7967$,

so $P(Z > 0.83) = 1 - 0.7967 =$ 0.2033

- Z Score is a standardized distance away from mean
- Probability is the area under the curve stopping at the Z Score threshold
- In example on left
  - Z Score is .83 (less than 1 std dev)
  - Area under the curve is .79
  - Given this data
    - 80% prob won't be stockout
    - 20% prob there will be

Covariance and Correlation

Sample Standard Error and Margin of Error

If a random sample of size *n* is taken from a (normal) population with mean μ and standard deviation σ, then the sampling distribution of the sample mean $\bar{x}$ has

mean: $E(\bar{x}) = \mu_{\bar{x}} = \mu$

standard error: $SE(\bar{x}) = \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

1. The best guess of μ is $\bar{x}$ and the reverse.
2. As *n* increases, estimates become more reliable.

	Survey 1	Survey 2	Survey 3	Survey 4	Survey 5	Survey 6
	4	7	6	6	4	4
	5	5	4	6	5	5
	7	4	5	6	7	7
	1	6	4	4	1	1
	2	1	4	1	1	2
AVG	3.8	4.6	4.6	4.6	3.6	3.8
Mean AVG	4.166667					

Only 1 Survey	
4	
5	
7	
1	
2	
AVG	3.8
SD	2.387467
SE	1.067708
.05 alpha	1.96
ME Low	1.707293
ME High	5.892707

95% Margin of Error = $1.96\frac{\sigma}{\sqrt{n}}$

T Test

1. Determine **null and alternative hypotheses.**	1. H_0: $\mu \le \$2000$ vs H_a: $\mu > \$2000$
2. Specify level of **significance** α.	2. $\alpha = 0.01$
3. Calculate **test statistic** value.	3. $z = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} = \frac{2157-2000}{581/\sqrt{115}} = 2.898$
4. Determine **critical value(s)**.	4. Reject H_0 if $z > z_{0.01} = 2.33$
5. Decide whether to **reject** H_0 and interpret the statistical result.	5. Since $z = 2.898 > 2.33$, we can **reject** H_0 and conclude the rental prices are high enough.

NOTE! SE is used when calculating Z Score for a sample

s/sqrt(n) = Standard Error

This distribution is null hypothesis distribution

This is where we found our sample to be

How is it that with our null hypothesis we were able to obtain a sample like we did?