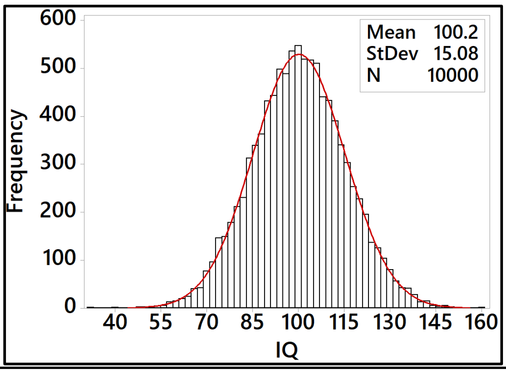
Normal Distribution

- Empirical Rule
- Normal Distribution
- Standard Normal Distribution
- Finding Normal Distribution Probabilities
- Finding Normal Distribution Percentiles
- Assessing Normality

University of Pittsburgh

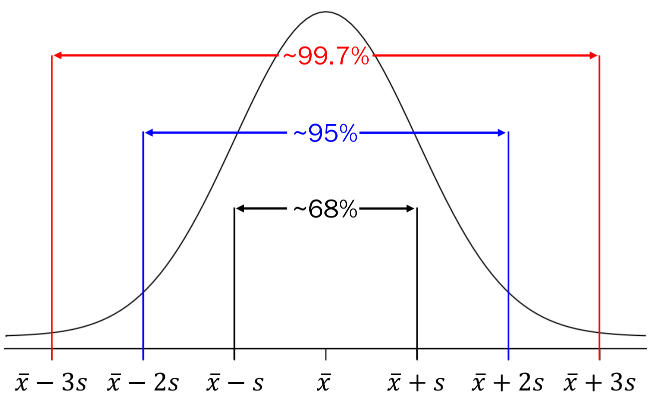
Example: Histogram to Probability Distribution

- **Scenario:** Random sample of 10,000 IQ scores returned a sample mean of $\bar{x} = 100.2$ and a sample standard deviation of $s = 15.08$.
- **Question:** What does the histogram below show us?
- **Answer:** IQ scores tend to follow a normal distribution with
  - Population mean: _________
  - Population standard deviation: _________
- **Takeaway:**
  - When a lot of quantitative data is available, a histogram approximates a _______________ distribution.
  - When the histogram is symmetric and unimodal, we can use the _________ **distribution.**



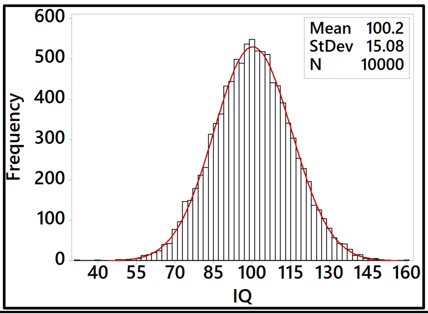
Review: Empirical Rule

- **Empirical Rule:** In data that is **approximately normal**, about:
  - 68% of observations are within 1 standard deviation of the mean
  - 95% of observations are within 2 standard deviations of the mean
  - 99.7% of observations are within 3 standard deviations of the mean



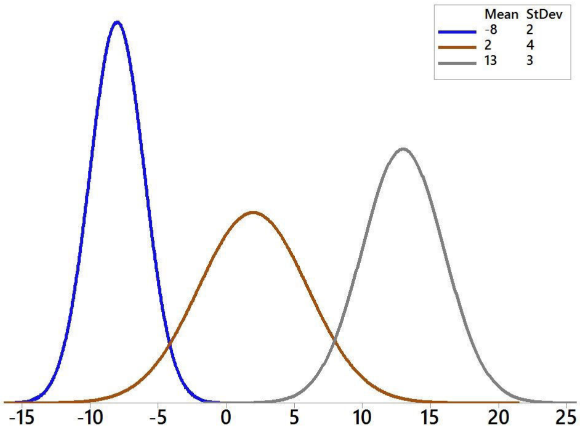
Example: Empirical Rule

- **Scenario:** IQ scores have a normal distribution with mean 100 and standard deviation 15.
- **Question:** What proportion of IQ scores are between 70 and 130?
- **Answer:** ______________
  - 70 and 130 are both ______________________ from the mean
- **Question:** What proportion of IQ scores are between 82 and 124?
- **Answer:** Empirical Rule not ______________
  - 82 and 124 lie _________________ standard deviations from the mean
  - **Best estimate:** ______________________
- **Takeaway:** Need a more ______________ to calculate probabilities



Normal Distribution

- **Normal distribution:** continuous probability distribution that describes data whose histogram is unimodal and symmetric
  - Defined by two parameters: Mean (μ) and standard deviation (σ)
  - Entire area under curve equals 1 → Areas are interpreted as probabilities



Observations

- Blue curve is tallest because it has the __________ spread
- Red curve is shortest because it has the __________ spread
- Grey curve is shifted farthest to the right because it has the ________________

Motivation: Standardization

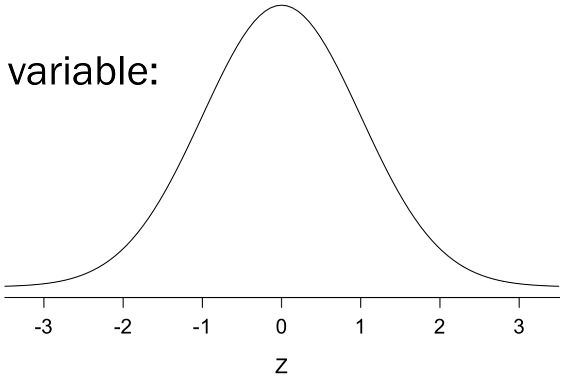
- **Scenario:** Amount of time college students study outside of class per week is normal with a mean of 24 hours and standard deviation 8 hours
- **Question:** How do we use the normal distribution to find the probability a student studying less than 18 hours per week?
- **Answer:** Need to ______________
  - No easy-to-use function or rule that calculates normal distribution probabilities
  - **Standardize** every normal random variable to find how many _________ __________ the observation is from the mean
  - Use the ______________________ to find probabilities

STANDARD NORMAL DISTRIBUTION

- **Standard normal distribution:** a special case of the normal distribution that has mean $\mu = 0$ and standard deviation $\sigma = 1$
  - Denoted by Z
  - Values referred to as Z-scores or Z-statistics
  - Probabilities found using a standard normal table
- To create a standard normal random variable:

$$Z = \frac{x - \mu}{\sigma} = \frac{\text{Observation} - \text{Mean}}{\text{Standard Deviation}}$$

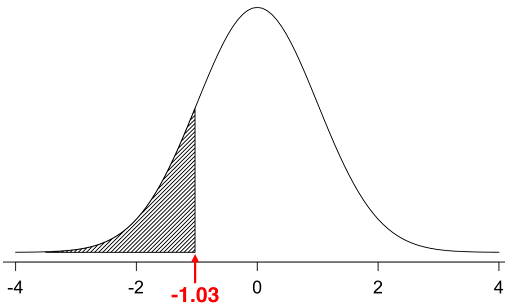
  - Creates a Z-score, which can be used to calculate probabilities



STANDARD NORMAL TABLE

- **Question:** How does the standard normal table work?
- **Answer:** Displays area to the **left** of the Z-score that is looked up

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641



- **Question:** What is the area to the left of -1.03?
- **Answer:** ________________

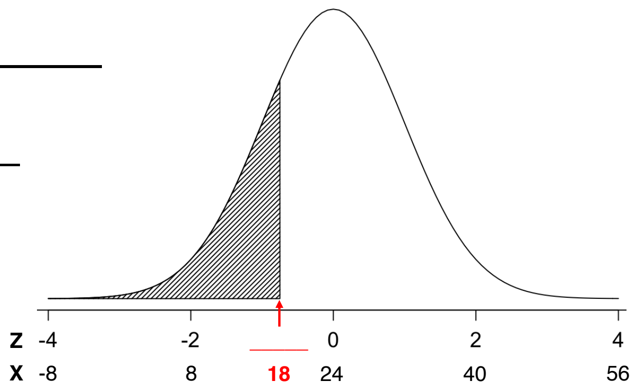
EXAMPLE: LOWER TAIL

- **Scenario:** Amount of time college students study outside of class per week is normal with a mean of 24 hours and standard deviation 8 hours
- **Question:** What is the probability a student studies less than 18 hours per week?
- **Answer:** Let $X =$ ____________________

__________ = ______________________

= ________________

= ________



Example: Upper Tail

- **Scenario:** Delayed flights at the Toronto airport are normal with a mean of 34 minutes and standard deviation of 11 minutes.
- **Question:** What is the probability a flight is delayed by more than 1 hour?
- **Answer:** Let $X =$ ____________________

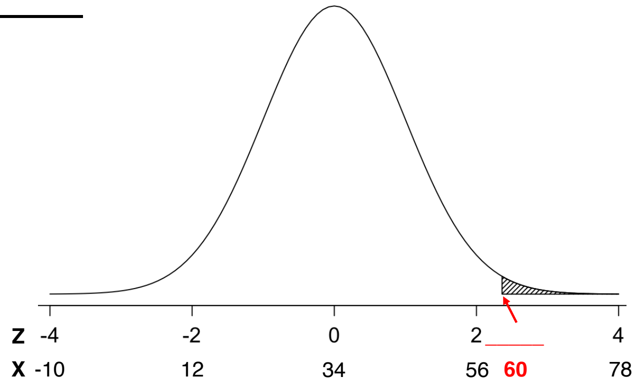
__________ = ________________

= ________________

= ________________

= ____________

= ________



Example: Middle of Distribution

- **Scenario:** IQ scores have a normal distribution with mean 100 and standard deviation 15.
- **Question:** What proportion of people have an IQ score between 82 and 124?
- **Answer:** Let $X =$ ____________

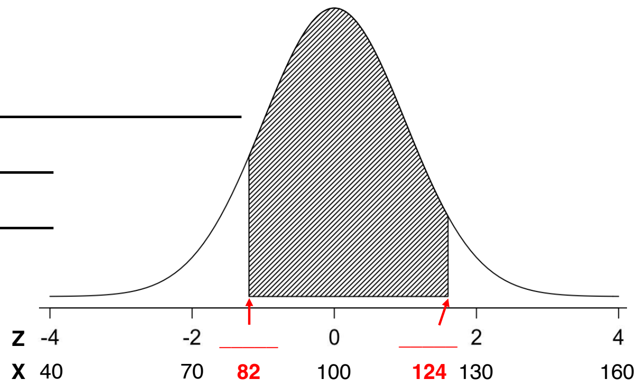
__________________ = ___________________________

= ________________________

= ________________________

= ________________

= ________



Example: Both Tails

- **Scenario:** High temperatures in Pittsburgh on June 7 historically have been normal with a mean of 78 degrees and SD of 7.30.
- **Question:** What is the probability the high temperature on one June 7 is at least 10 degrees away from the mean?
- **Answer:** Let $X =$ ___________________________

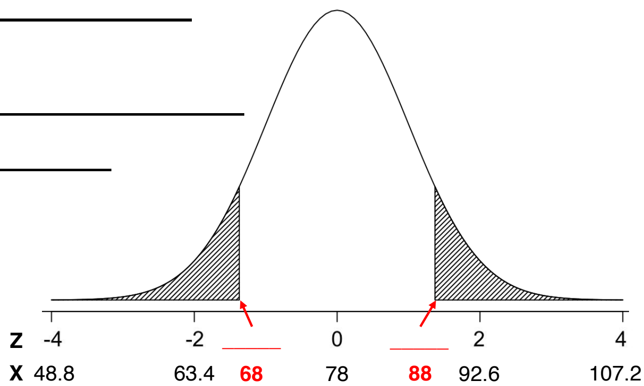
___________________ = _________________________

= ___________________________

= ________________

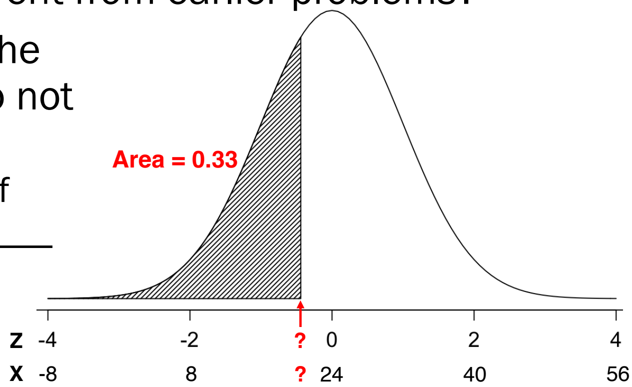
= _____________

= ________



MOTIVATION: NORMAL DISTRIBUTION PERCENTILES

- **Scenario:** Amount of time college students study outside of class per week is normal with a mean of 24 hours and standard deviation 8 hours. Want to know how many hours the least studious 33% of students study.
- **Question:** What makes this different from earlier problems?
- **Answer:** Know the ______ under the normal distribution curve, but do not know the ____________________
  - Previously, we knew the number of hours and calculated the __________
- **Solution:** Calculate _________ ____________________________

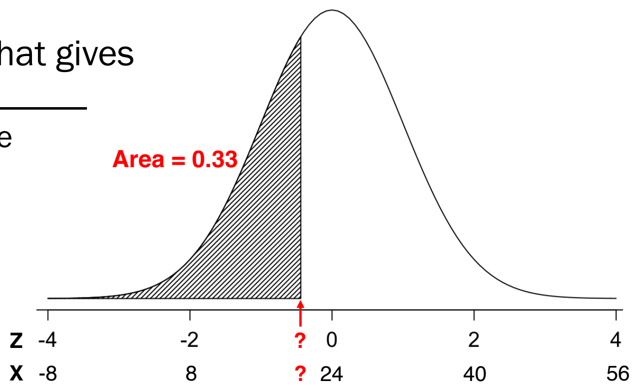


NORMAL DISTRIBUTION PERCENTILES

- To calculate normal distribution percentiles, we need to answer two questions:
  1. What Z-score corresponds to the desired area in the tail?
     - **To solve:** Find the percentile **inside** the standard normal table and solve for the Z-score by moving left and up
  2. Given the Z-score from Step 1, what value is this many standard deviations above/below the mean
     - **To solve:** Use the mean, standard deviation, and Z-score from Step 1 to unstandardize $Z = \frac{X-\mu}{\sigma}$ and solve for X

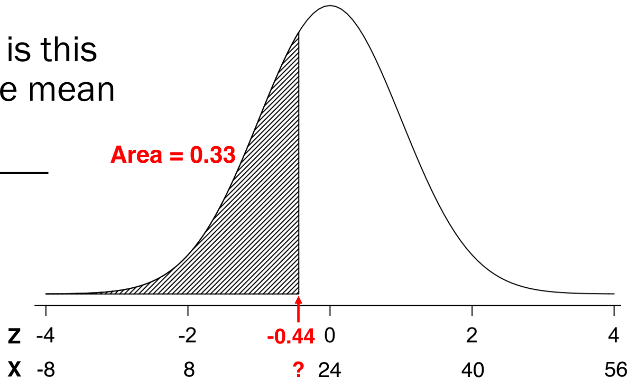
EXAMPLE: LOWER PERCENTILES

- **Scenario:** Amount of time college students study outside of class per week is normal with a mean of 24 hours and standard deviation 8 hours
- **Question:** What is the 33rd percentile of all weekly study times?
- **Answer:**
  - First need to know the ________ that gives us _______________ in the ___________
    - Find the _____________ inside the standard normal table
    - Solve for the _______
  - **33rd percentile:** ______________



EXAMPLE: LOWER PERCENTILES

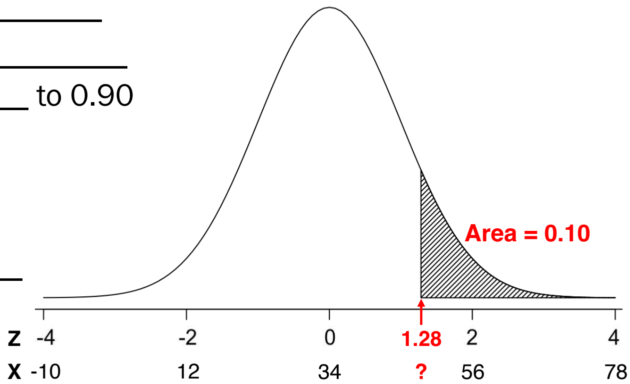
- **Scenario:** Amount of time college students study outside of class per week is normal with a mean of 24 hours and standard deviation 8 hours
- **Question:** What is the 33rd percentile of all weekly study times?
- **Answer (cont.):**
  - Second, calculate the ______ that is this many _________________ from the mean
  - Set up the Z-score: _______________
  - Solve for the _______________
    - $X =$ ___________ $=$ ______
  - 33% of students study _______ _______________ per week



EXAMPLE: UPPER PERCENTILES

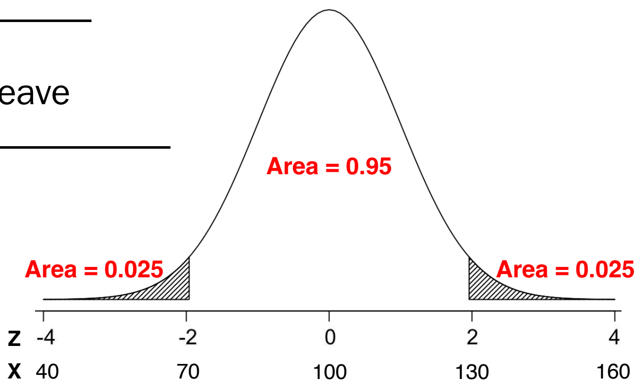
- **Scenario:** Delayed flights at the Toronto airport are normal with a mean of 34 minutes and standard deviation of 11 minutes.
- **Question:** How long are the longest 10% of delays?
- **Answer:**
  - "Longest 10%" means _______________
  - **Problem:** 0.90 is not __________________
    - **Solution:** Find the _______________ to 0.90
  - 90th percentile: _____________
  - Z-score: ______________

    $X =$ ___________ $=$ _____
  - 10% of flights will have a delay of at least _______________



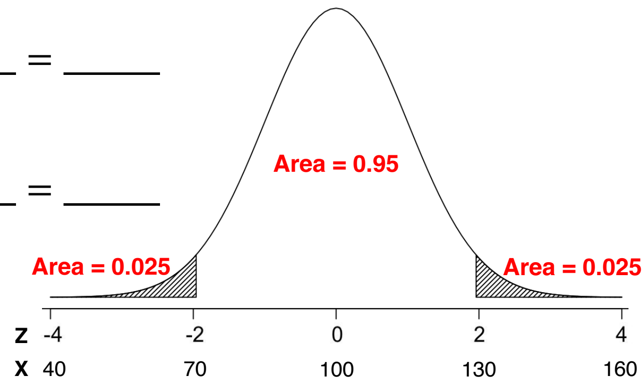
EXAMPLE: MIDDLE OF DISTRIBUTION

- **Scenario:** IQ scores have a normal distribution with mean 100 and standard deviation 15.
- **Question:** What values yield the middle 95% of all IQ scores?
- **Answer:**
  - Need _____ percentiles – one in ______ tail and one in ______ tail
  - "Middle 95%" means we need to leave out the ____________ and the ____________
    - 2.5th percentile: ____________
    - 97.5th percentile: ____________
  - **Note:** Because the same area is left out of each tail, the Z-scores will have the same ____________.



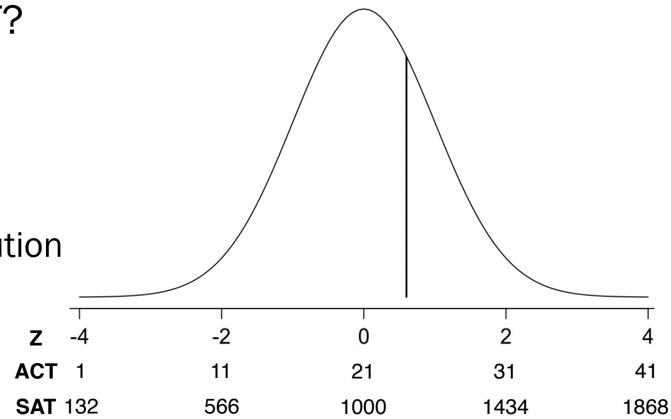
Example: Middle of Distribution

- **Scenario:** IQ scores have a normal distribution with mean 100 and standard deviation 15.
- **Question:** What values yield the middle 95% of all IQ scores?
- **Answer:**
  - Lower bound: ______________
    - $X =$ ____________ $=$ _____
  - Upper bound: ______________
    - $X =$ ____________ $=$ _____
  - 95% of all IQ scores lie between ______ and ______
    - Empirical Rule was __________



Example: Converting Between Two Distributions

- **Scenario:** ACT scores are normal with a mean of 21 and a standard deviation of 5. SAT scores are also normal with a mean of 1000 and a standard deviation of 217.
- **Question:** If a student scored 24 on the ACT, what would the equivalent score be on the SAT?
- **Answer:**
  - Find the Z-score for the _____
    - $Z =$ ________ $=$ _____
  - ____________ using the distribution of the _____
    - $X =$ ____________ $=$ _____

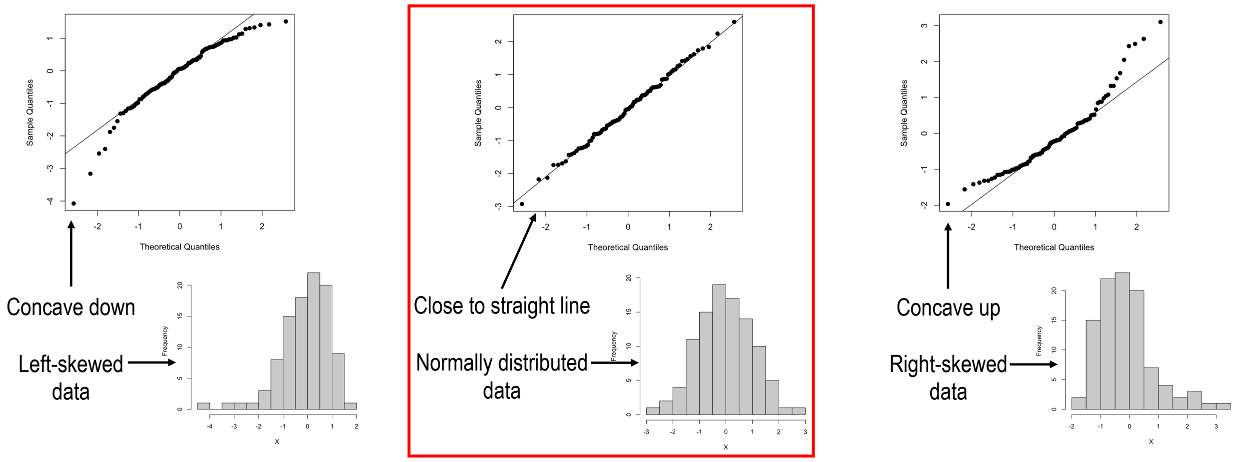


Looking Ahead to Inference: Assessing Normality

- In **probability**, we often know if a variable is normally distributed based on having a large amount of historic data.
  - Allows us to calculate probabilities and percentiles immediately.
- In **inference**, we only work with samples. Before analyzing the data, it is often important to determine if a variable is normal by analyzing the sample.
- The normality of a sample can be assessed using a **normal probability plot.**
  - Histograms also work but may not paint a clear picture for small samples.

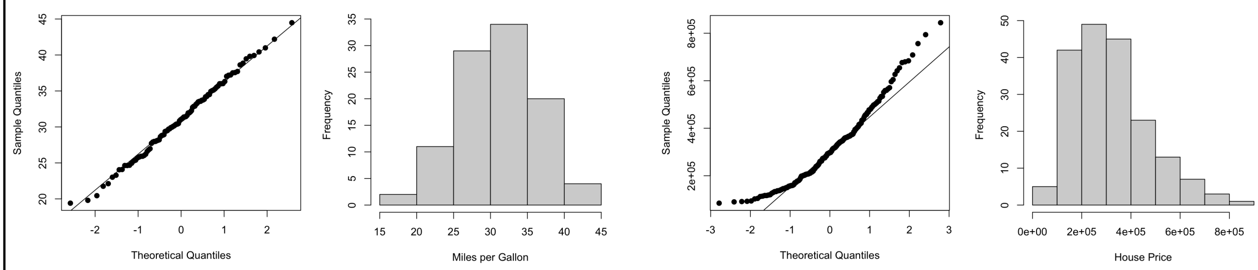
Normal Probability Plot

- **Normal probability plot:** a plot of the observations' Z-scores in increasing order (Y-axis) against the Z-score the observation would have if the data were perfectly normally distributed (X-axis)



Example: Assessing Normality

- **Scenario:** Normal probability plot for the fuel efficiencies (in MPG) for 100 cars on the left and 188 house prices on the right
- **Question:** What can be said about the shape of the distributions?



- **Answer:**
  - **Fuel Efficiencies:** ____________________________ – plot is ________________________
  - **House Prices:** __________________ – plot is ________________