

Problem 1)

age: 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70

- a) Bin 1: $(13+15+16)/3 = 14.67$, Bin 2: $(16+19+20)/3 = 18.333$, Bin 3: $(20+21+22)/3 = 21$, Bin 4: $(22+25+25)/3 = 24$, bin 5: $(25+25+30)/3 = 26.67$, bin 6: $(33+33+35)/3 = 33.67$, bin 7: **35**, bin 8: $(36+40+45)/3 = 40.33$, bin 9: $(46+52+70)/3 = 56$. **Smoothing the data gives us fewer numbers to work with and reduces the amount of small variations**
- b) Q1: 25, Q3: 35, IQR: $35-25 = 10$, lower bound: $25 - (1.5*10) = 10$, upperbound: $35 + (1.5*10) = 50$, **52 and 70 are outliers**
- c) $(35 - 13) / (70-13) = .386$
- d) Mean = $(13 + 15 + 16 + 16 + 19 + 20 + 20 + 21 + 22 + 22 + 25 + 25 + 25 + 25 + 30 + 33 + 33 + 35 + 35 + 35 + 35 + 36 + 40 + 45 + 46 + 52 + 70) / 27 = 29.963$, standard deviation: 11.84, $Z = (35-29.963) / 11.84 = .43$
- e) $70 / 10^j < 1$, $70 / 10^2 = 0.7$, $35 / 10^2 = 0.35$

Problem 3)

Entropy:

PJunior = 113/165, Psenior = 52/165

$H(S) = -(.685 \log_2 (.685) + .315 \log_2 (.315)) = -((.685 * -.546) + (.315 * -1.667)) = -(-.374 - .525) = .899$

Each Department:

Sales = $-(80/110 \log_2 80/110 + 30/110 \log_2 30/110)$

Systems = $-(26/37 \log_2 26/37 + 11/37 \log_2 11/37)$

Marketing = $-(4/14 \log_2 4/14 + 10/14 \log_2 10/14)$

Secretary = $-(6/10 \log_2 6/10 + 4/10 \log_2 4/10)$

Department = $110/165 * .845 + 31/165 * .826 + 14/165 * .863 + 10/165 * .971 = .85$

Gain department = $.899 - .85 = .049$

Age:

31_35: 44 juniors, 35 seniors

$H(31_35) = -(44/79 \log_2 44/79 + 35/79 \log_2 35/79) = .988$

Weighted average: $H(\text{age}) = 79/165 * .988 + 0 + 0 = .473$

Gain(age) = $.899 - .473 = .426$

Salary:

46k_50k 23 juniors, 40 seniors

$H(46k_50K) = -(23/63 \log_2 23/63 + 40/63 \log_2 40/63) = .946$

$H(\text{salary}) = 63/165 * .946 + 0 + 0 + 0 = .361$

Gain salary = $.899 - .361 = .538$

First split: salary, second split: department

Problem 4.

If salary is 26_30k, 31k_35k, 36k_40k, 41K_45K, 66_70k Then

Junior 26_30k, 31k_35k, 36k_40k, 41K_45K

Senior 36k_40k, 66_70k

If salary = 46k_50k Then

If department sales, marketing then senior

If department systems then junior