

Statistics 528: Homework 2 Solutions

1.41 There are several gaps in the data, as can be seen from the histogram.

Minitab Result:

Min	Q1	Med	Q3	Max
2508	13278	17022	23568	26241

Manual Result:

Min	Q1	Med	Q3	Max
2508	13385	17022.5	23552	26241

We cannot find out the overall shape or distribution from the five-number summary. From histogram, we can see there may be 3 modes, also a gap between \$6,000 & \$9,000, which couldn't be seen from summaries.

1.45

(a)

Stem-and-Leaf Display: Fees

Stem-and-leaf of Fees N = 50
Leaf Unit = 1.0

```

 2   0 89
 7   1 00234
15   1 55558899
(28) 2 0000000000000111112222222223
 7   2 59
 5   3 0
 4   3 5
 3   4 00
 1   4
 1   5 0
    
```

\$20 is the most common value.

(b)

IQR=4, and Q3=22, so anything over 28 is a possible outlier. There are several values over \$30.

1.50

Without the outlier, the mean is 137.59 (compared to before 141.06). The median is 137 (compared to 138.5). The mean changed by about 3.5, while the median changed by only 1.5 as a result of dropping the outlier.

1.58

(a) Back-to-back stemplot:

Leaf unit: 10 minutes

```

          433330      0
        99997666      0    69
          2222222      1    12222222
            855      1    555578888888888888
           4400      2    0444
                2
                3
                3    6
    
```

Symmetric, bell-shaped distributions without outliers are better summarized using \bar{x} and s .

(b)

Descriptive Statistics: Minutes by Gender: (Including all data)

Variable	Gender	N	Mean	Median	TrMean	StDev
Minutes	F	30	165.2	175.0	161.7	56.5
	M	30	117.2	120.0	113.3	74.2
Variable	Gender	SE Mean	Minimum	Maximum	Q1	Q3
Minutes	F	10.3	60.0	360.0	120.0	180.0
	M	13.6	0.0	300.0	60.0	157.5

The question asks if it appears to contain a high outlier. So we only need the upper limit.

Female: upper limit = $Q3 + 1.5 * IQR = 180 + 1.5 * (180 - 120) = 270$

Male: upper limit = $Q3 + 1.5 * IQR = 157.5 + 1.5 * (157.5 - 60) = 303.75$

In the female group, 360 is flagged as suspicious by 1.5*IQR criterion. In the male group, 300 is marginally within 1.5*IQR.

Descriptive Statistics: Minutes by Gender (excluding the highest outlier in each group)

Variable	Gender	N	Mean	Median	TrMean	StDev
Minutes	F	29	158.45	170.00	159.07	43.65
	M	29	110.9	120.0	110.2	66.9
Variable	Gender	SE Mean	Minimum	Maximum	Q1	Q3
Minutes	F	8.11	60.00	240.00	120.00	180.00
	M	12.4	0.0	240.0	60.0	150.0

For female students, removing the outlier 360 resulted in the change of the mean from 165.2 to 158.45 and standard deviation from 56.5 to 43.65. Likewise, for male students

the mean decreases by 6.3 from 117.2 to 110.9 and the standard deviation decreases by 7.3 from 74.2 to 66.9 when we remove the outlier.

1.72

(a) $a=0$, $b=1/0.62=1.613$; $65\text{mph}=104.8\text{km/h}$

$X_{\text{new}}=a+bX$, when $X=0$, $X_{\text{new}}=0$, it means $a=0$. $X_{\text{new}}=bX$

When $X=0.62$, $X_{\text{new}}=1$, it means $b=1/0.62=1.613$. $X_{\text{new}}=1.613X$

When $X=65$, $X_{\text{new}}=1.613*65=104.8\text{km/h}$

(b) $a=0$, $b=746$; $140\text{hp}=104,440\text{watts}$

$X_{\text{new}}=a+bX$, when $X=0$, $X_{\text{new}}=0$, it means $a=0$. $X_{\text{new}}=bX$

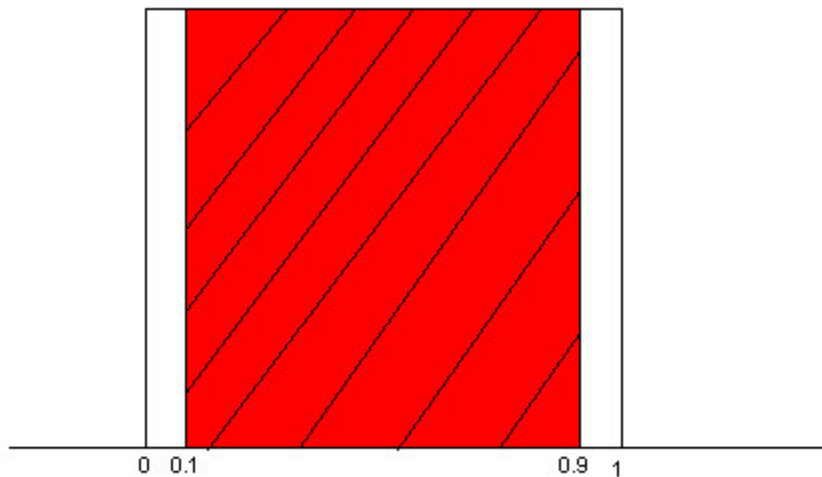
When $X=1$, $X_{\text{new}}=746$, it means $b=746$. $X_{\text{new}}=746X$

When $X=140$, $X_{\text{new}}=140*746=104,440\text{watts}$

1.78

(a) The curve forms a box that is 1 unit wide and 1 unit high. Therefore, the area is $1*1=1$

(b) The shaded area is 0.8 wide and 1 unit high. Therefore, the area is 0.8.



1.86

For Eleanor, $z=(680-500)/100=1.8$.

For Gerald, $z=(27-18)/6=1.5$.

Eleanor's score is higher.

1.88

(a) $P(Z<2.85)=0.9978$

(b) $P(Z>2.85)=1-P(Z<2.85)=1-0.9978=0.0022$

(c) $P(Z>-1.66)=P(Z<1.66)=0.9515$ by the symmetry of the standard normal distribution

(d) $P(-1.66<Z<2.85)=P(Z<2.85)-P(Z<-1.66)=0.9493$

1.98

(a) Let X denote the yearly return, can be approximated by the $N(13, 17)$. By the 68-95-99 criterion, approximately 95% of years lie within the range of 2 standard deviation from the mean. Therefore, from $13 - 2 * 17 = -21\%$ to $13 + 2 * 17 = 47\%$

(b) The percent of years when the yearly return is below 0 is

$$P(X < 0) = P((X - 13) / 17 < (0 - 13) / 17) = P(Z < -0.765) = 0.22$$

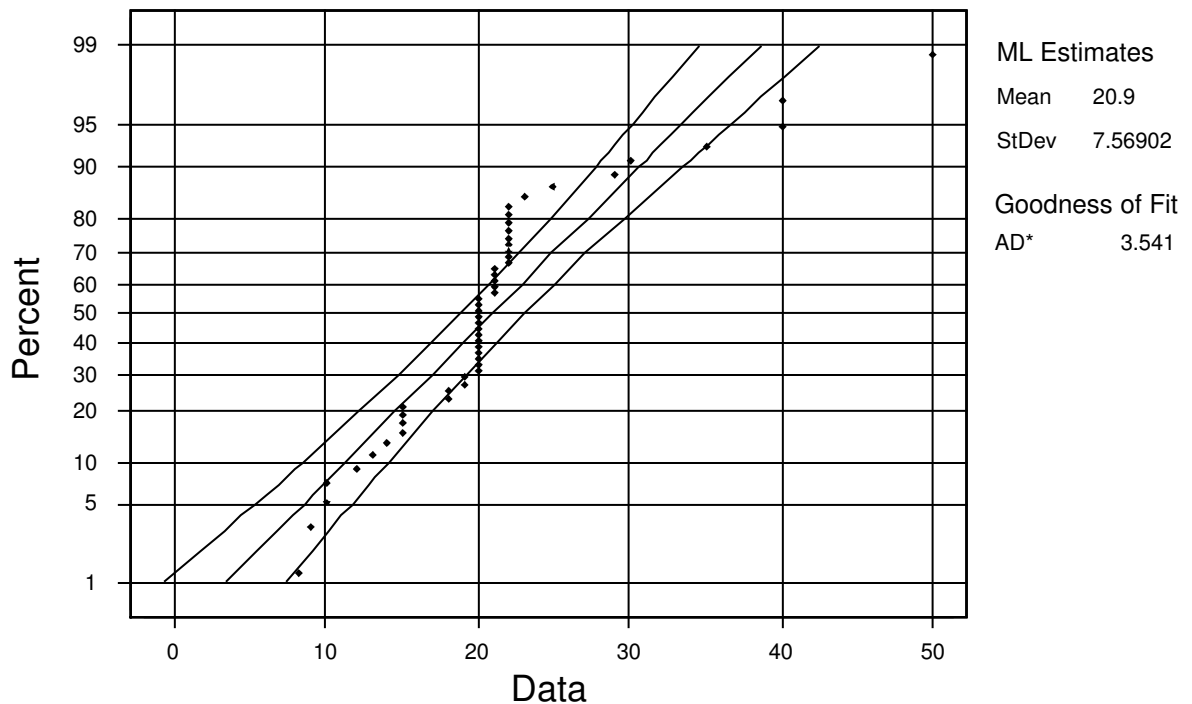
(c) The percent of years when the yearly return is above 25% is $P(X > 25) = P((X - 13) / 17 > (25 - 13) / 17) = P(Z > 0.706) = 0.22$

1.112

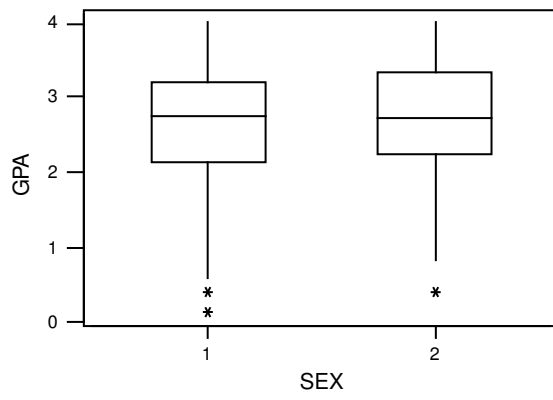
On the right end, the data are higher than predicted for a normal curve. Note that the Minitab uses the data values for the x-axis and the normal percentiles for the y-axis. Therefore, when the data on the right end fall below the straight line, it indicates a heavier tail than normal distribution or skewness to the right. (Note: this is different from the text book since normal quantile plots in the text use the data values for the y-axis.) Likewise, when data on the left end fall above the straight line, it indicates skewness to the left.

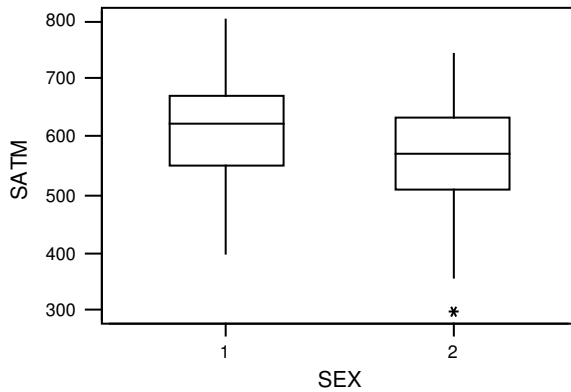
Normal Probability Plot for Fees

ML Estimates - 95% CI



1.140

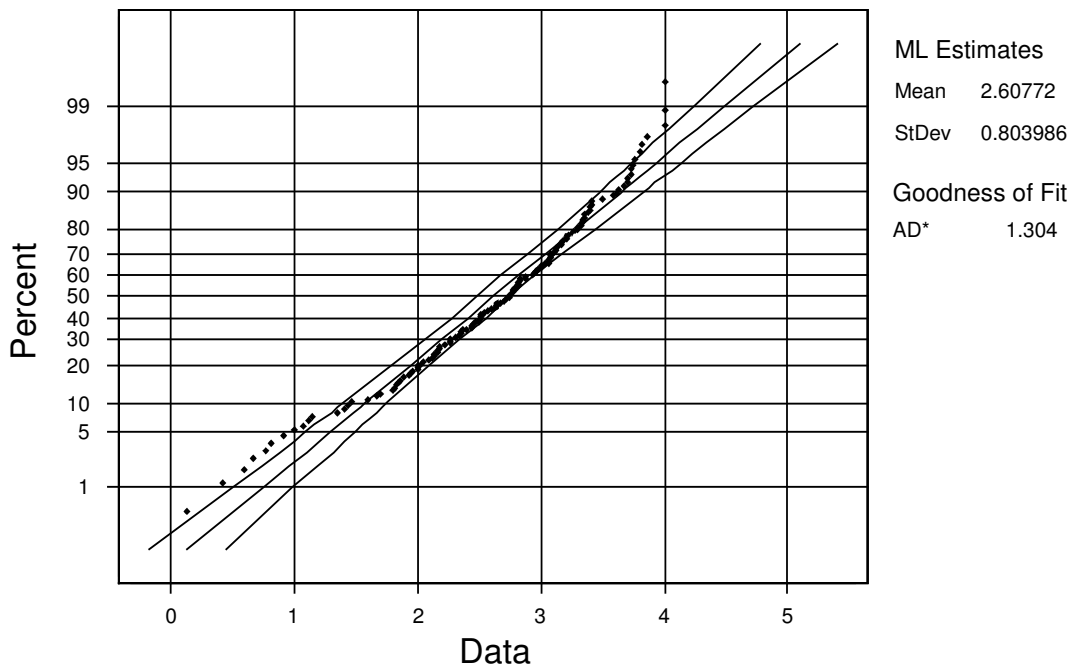




From the plots, we see that the medians of GPA are fairly similar for both groups. The median of men's SATM is higher than that of women's. There seems no difference in the spread of two distributions for men and women.

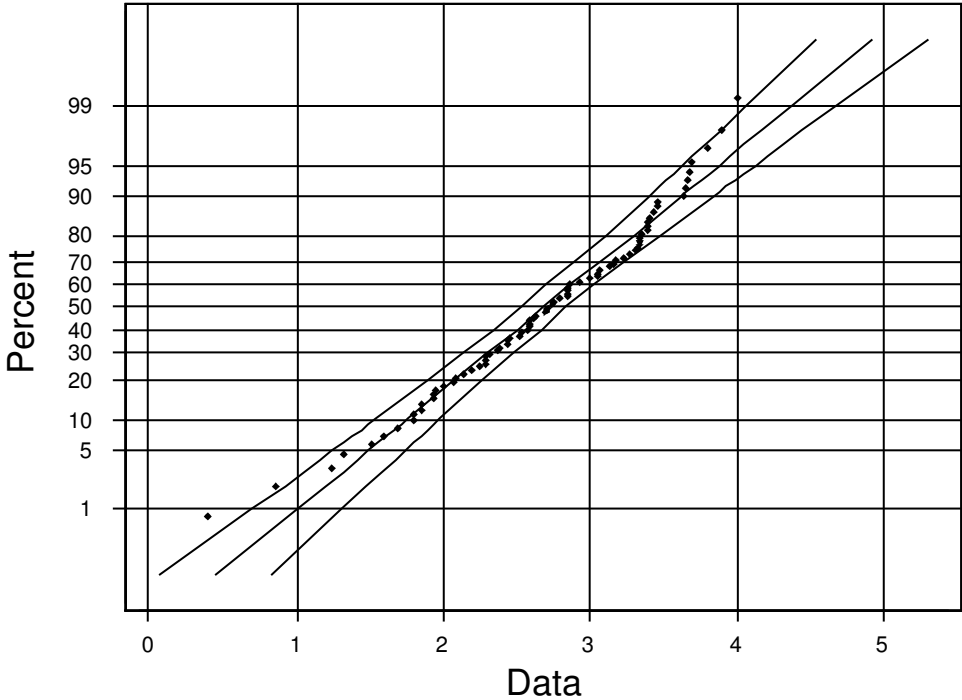
Among the four distributions below, the distribution of male students' SATM has a normal quantile plot with most of the points falling approximately on the straight line in the middle, and within the band.

Normal Probability Plot for GPA of Men
ML Estimates - 95% CI



Normal Probability Plot for GPA of Women

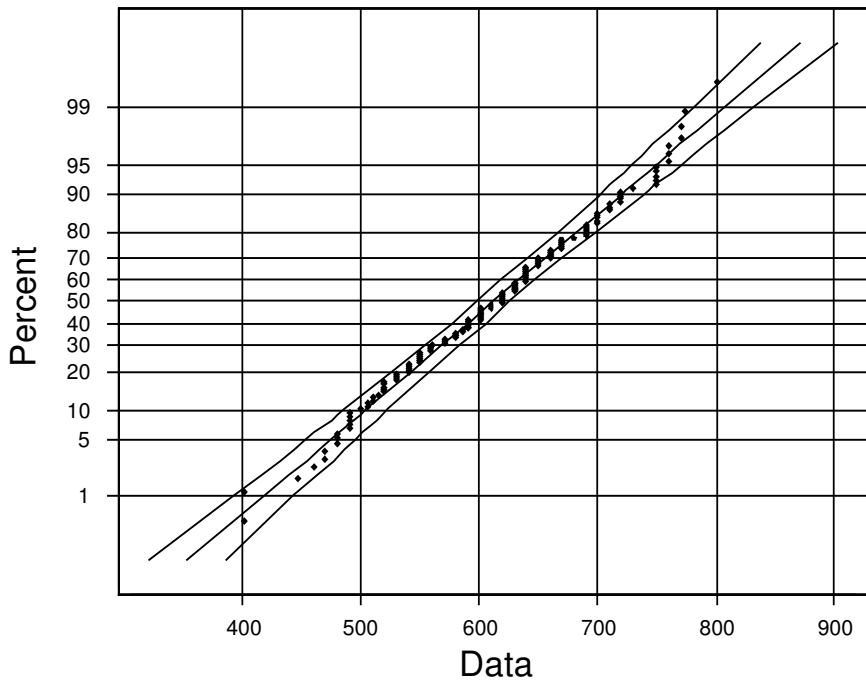
ML Estimates - 95% CI



ML Estimates
Mean 2.68570
StDev 0.724130
Goodness of Fit
AD* 0.582

Normal Probability Plot for SATM of Men

ML Estimates - 95% CI



ML Estimates

Mean 611.772

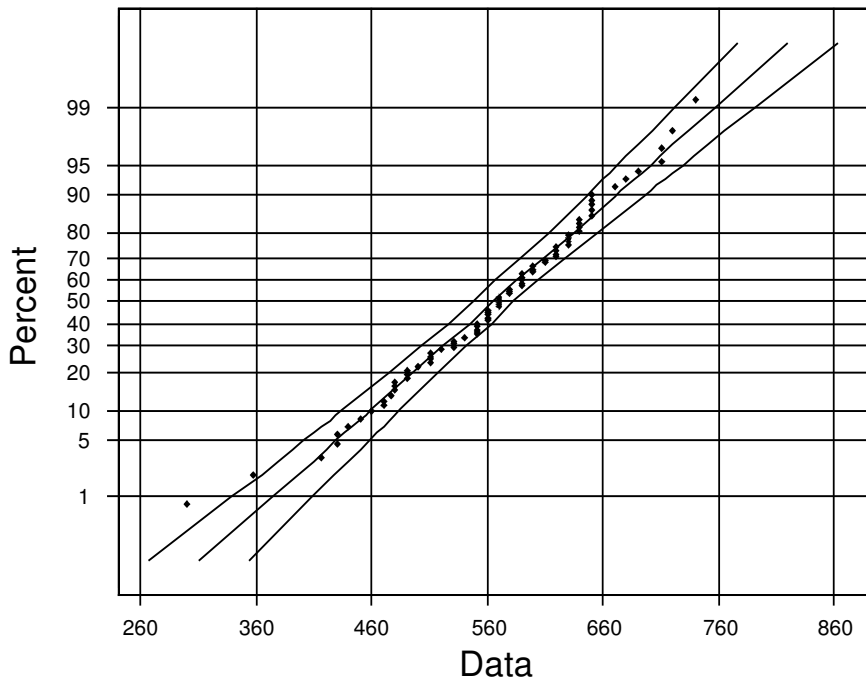
StDev 83.7303

Goodness of Fit

AD* 0.406

Normal Probability Plot for SATM of Women

ML Estimates - 95% CI



ML Estimates

Mean 565.025

StDev 82.4028

Goodness of Fit

AD* 0.464