

Solutions to sheet 1

I discovered how to calculate with Minitab by first using the menus with the command language enabled. That way Minitab would put a copy of the commands into the top window. After a while I learned what to type directly into the command window.

(1.1)

The Yale grades data were in a worksheet with 39 rows, and columns including 'final' and 'classwork'. Calculate standardized versions.

```
MTB > name c6 = 'final.std'
MTB > Center 'final' c6
MTB > name c7 'cwork.std'
MTB > center 'classwork' c7
```

Get IQR for final, storing in c8 (use the **Stat>Basic Statistics>Store Descriptive Statistics** menu)

```
MTB > Name c8 = 'IQR1'
MTB > Statistics 'final';
SUBC> IQRRange 'IQR1'.
```

Divide final by its IQR, and save.

```
MTB > let c9 = final/iqr1
name c9 'final.iqr'
```

Similar calculation for classwork:

```
MTB > Name c10 = 'IQR2'
MTB > Statistics 'classwork';
SUBC> IQRRange 'IQR2'.
MTB > let c11 = classwork/iqr2
MTB > name c11 'cwork.iqr'
```

I wanted the ranks to run from 1 (first in class) to 39:

```
MTB > let c12 = 40-rank(final+classwork)
MTB > name c12 = 'rank1'
MTB > let c13 = 40 - rank(final.std+cwork.std)
MTB > name c13 = 'rank2'
MTB > let c14 = 40 - rank(final.iqr+cwork.iqr)
MTB > name c14 = 'rank3'
```

Sort rows so that the column for rank1 (that is, by sum of raw scores) goes from 1 to 39. Note the decimal fractions denoting tied scores.

```
MTB > Sort 'final' 'classwork' 'rank1' 'rank2' 'rank3' c15-c19;
SUBC> By 'rank1'.
```

Rename the columns c15-c19 as `_final_`, `_cwork_`, `_rank1_`, `_rank2_`, and `_rank3_` then print the ranks according to the three methods, together with the scores that got them there:

```
MTB > print c15-c19
```

Row	<code>_final_</code>	<code>_cwork_</code>	<code>_rank1_</code>	<code>_rank2_</code>	<code>_rank3_</code>
1	73	596	1.0	2	2
2	76	591	2.0	1	1
3	70	592	3.0	3	3
4	65	596	4.0	5	5
5	69	587	5.0	4	4

6	57	588	6.0	13	11
7	53	587	7.0	15	15
8	63	573	8.5	9	8
9	67	569	8.5	6	6
10	59	575	10.5	14	12
11	63	571	10.5	11	9
12	66	561	12.0	8	10
13	68	558	13.0	7	7
14	64	558	14.0	12	13
15	55	560	15.5	16	16
16	67	548	15.5	10	14
17	51	558	17.0	19	17
18	44	561	18.0	22	18
19	38	548	19.0	24	24
20	51	531	20.0	20	20
21	47	526	21.0	23	22
22	28	535	22.5	30	28
23	62	501	22.5	17	19
24	38	516	24.0	26	26
25	39	509	25.0	27	27
26	45	500	26.0	25	25
27	59	483	27.0	21	23
28	65	475	28.0	18	21
29	36	497	29.0	28	29
30	35	495	30.0	29	30
31	21	504	31.0	35	32
32	38	470	32.0	31	31
33	27	457	33.0	36	35
34	31	452	34.0	34	34
35	44	421	35.0	32	33
36	35	390	36.0	37	36
37	57	297	37.0	33	37
38	32	310	38.0	38	38
39	31	176	39.0	39	39

The rankings for the two standardized methods are mostly in agreement. Use of IQR instead of standard deviation to rescale has the advantage that extreme outliers can't have influence on the measure of spread in the scores. I think it would be unfair if a few extreme cases could inflate a measure of spread, and thereby reduce the influence of a score on the ultimate grade for everyone.

The few students (such as rows 6, 7, and 22) who dropped substantially from their raw rankings performed far worse on the final exam than others who had similar rankings by method 1. What caused the rises in rank?

To me it seems fair that 10 points on classwork should count for less than 10 points on the final exam, for the raw scores as they stand. Mere addition of the two raw scores would give equal weight to 10 points in either score. Moreover the total number of points for the final exam was quite arbitrary. If I had decided to mark the final out of 8000, and then I had just combined raw scores, I could have wiped out the effect of classwork on the rankings. Some form of scaling seems essential if arbitrary decisions about scoring are not to be transmitted into the rankings.

(1.2)

There are several ways to carry out the necessary calculations. It is probably easier to have a calculator handy to work out some of the numbers directly. If you are more adventurous, you can use Minitab as your calculator, as shown below.

First convert MAL to mph.

Then create a new column with a 1 whenever the MAL.mph lies in the strong breezy range, 0 otherwise. Take the mean of the new column to get the proportion of 1's, that is, the fraction of strongly breezy days.

```
MTB > Let c7 = Mal.mph >= 25 and MAL.mph <= 31
MTB > let K3 = mean(c7)
MTB > print k3
K3      0.118953 <- fraction of strong breezy days
```

Or you can do it all in one step:

```
MTB > let k4 = mean(Mal.mph >= 25 and MAL.mph <= 31)
MTB > print k4
K4      0.118953
```

Use Minitab as a calculator:

```
MTB > let k5 = mean(sqrtMAL)
MTB > let k6 = stdev(sqrtMAL)
MTB > name k6 'sqrtMAL.stddev'
MTB > name k5 'sqrtMAL.mean'
```

Or:

```
MTB > Name c8 = 'Mean1' c9 = 'StDev1'
MTB > Statistics 'sqrtMAL';
SUBC> Mean 'Mean1';
SUBC> StDeviation 'StDev1'.
```

Enter values 25 and 31 into column c10, and name it beaufort. Then find corresponding range for square roots:

```
MTB > let c11 = sqrt(beaufort)
MTB > name c11 'sqrtBeau'
```

Use the **Calc > Probability Distributions > Normal** menu to find areas under normal curve to the left of sqrt(25) and sqrt(31). Save results in c12, then take difference to get area under normal curve between sqrt(25) and sqrt(31).

```
MTB > CDF 'sqrtBeau' c12;
SUBC> Normal 'sqrtMAL.mean' 'sqrtMAL.stddev'.
MTB > let k9 = c12(2) - c12(1)
MTB > print k9
K9      0.113989 <- area under approximating curve corresponding to strong breeze
```

Alternatively, you could calculate $(\text{sqrt}(31) - \text{mean}(\text{sqrtMAL})) / \text{stdev}(\text{sqrtMAL})$ and $(\text{sqrt}(25) - \text{mean}(\text{sqrtMAL})) / \text{stdev}(\text{sqrtMAL})$, evaluate the **standard** normal cdf at those two limits, then take the difference. (That is the way one does it with tables.)

In summary,

- from K4, proportion of strong breezy days at MAL is 11.9%
- from K9 (using normal approximation to histogram for sqrtMAL) proportion of strong breezy days should be about 11.4%

The point of the second calculation was to check how well the approximation works.