- 1. when we count something the tally is the frequency, f. Relative frequency is the proportion of times it occurs or f/n
- 2. \bar{x} is the sample mean, *n* the sample size, Σ the summation symbol, *s* the sample standard deviation, and Q_1 the first quartile.
- 3. A sample is the data we have. The population is the set of all possible values for a point in the sample. A statistic summarizes a sample, a parameter summarizes the population.

the random variable makes this more clear. a (random) sample is a collection of independent identically distributed random variables that have been observed. the population is the common distribution.

- 4. year in school: categorical, number of credits: discrete numeric, gpa: continuous numeric; time to commute: continuous numeric.
- 5. relative frequencies are the frequencies divided by n: 10/30, 15/30, 5/30. a barplot would look like (how does this differ from a histogram?):

echo=false, fig=true car = c(10,15,5) names(car) = c("Domestic","Imported","No car") barplot(car)

6. The 8 numbers are

1 1 1 2 3 5 8 13

The median is between 4th and 5th after sorting: M = 2.5. Q_1 is the median of 1,1,2,3 or 1.5. Q_3 is the median of 3,5,8,13 or 6.5 The min is 1, the max is 13.

THe IQR is $Q_3 - Q_1 = 6.5 - 1.5 = 5$ A boxplot looks like



7. The mean is directly: (or use mean())

> x = c(2, 3, 3, 5, 7)
> sum(x)/5
[1] 4
The standard deviation is found with

```
> x - mean(x)
[1] -2 -1 -1 1 3
> (x - mean(x))^2
[1] 4 1 1 1 9
> sum((x - mean(x))^2)
[1] 16
> sqrt((1/4) * sum((x - mean(x))^2))
[1] 2
```

- 8. (a) The min is between 0 and 100, the max between 600 and 700. YOu could say conservatively the range is between 0 and 700.
 - (b) The shape: unimodal, skewed right
 - (c) The mean is a bit more than the median, as the data is skewed right. I would guess by looking for a balancing point, that the mean is around 150, by looking to split the are in half that the median is around 80, and Q_1 is 40, and Q_3 is 225 so the IQR is about 225 40 = 185.
 - (d) There are about 70 numbers less than 100 and 10 more than 200.
- 9. (a) The median is marked by the solid line in the middle of the box. The "now" category has the smallest value. (Smoking reduces median birthweight!)
 - (b) The spread is read from the IQR, or the length of the box. In this case the NA category has the most spread.
 - (c) The tails are long if there are a lot of outliers stretching out. Only "never" has a long tail.
 - (d) Skew is shown by the boxplots being not symmetric about the median. In this case the data is not skewed.
 - (e) For "never" we have 55, 115, 125,130 adn 178 (or so) for the f-number summary.
- 10. The z-score is simply the number of standard deviations from the mean.
 - (a) For 82 it is

> (82 - 72.57)/8.638

[1] 1.092

(b) Is the z score fo 59 less than -2? Well 59 - 72 is -13 which is not 2 standard deviations, so the answer is NO. The actual z-score is

> (59 - 72.57)/8.638

[1] -1.571

- (c) Skip percentile rank, I copied this from a previous review sheet. We didn't cover it.
- 11. (a) The outcome space is the set of all outcomes. There are $2 \cdot 2 \cdot 2$ outcomes (why)

HHH HHT HTH HTT THH THT TTH TTT

(b) 2 heads is the event

HHT HTH THH

As there are 3 outcomes, and 8 equally likely outcomes in total the probability is 3/8.

(c) Exactly 1 tails consists of

ТНН НТН ННТ

again there are 3 outcomes, so this event has probability 3/8.

(d) More heads than tails is this set of outcomes:

ННН ННТ НТН ТНН

So the probability is 4/8 = 1/2. ANother way of doing this is realizing that this is 3 heads or 2 heads and adding 1/8 and 3/8 to get the answer.

- (e) Atleast on tail, is the complement of 0 tails which is the single outcome HHH. So the probability of atleast one tail is 1 minus the probability of 0 tails or 1 1/8 = 7/8.
- 12. We use the multiplication rule here:

> 6 * 2 * 2 * 3 * 3 * 2 * 3

[1] 1296

13. These aren't "cominbations" in the classes sense. We do this with the multiplication rule:

> 40 * 39 * 39

[1] 60840

(Why 39 for the last two?)

14. This combination lock really uses combinations: order isn't important. The answer is $\binom{n}{k}$:

> choose(10, 4)

[1] 210

> choose(10, 3)
[1] 120
> choose(10, 5)
[1] 252