Law School Admissions

As we would expect, accepted students have higher LSAT scores and grade point averages.

```
# A tibble: 2 x 4
  accepted total avg_lsat avg_gpa
  <lgl>    <int>    <dbl>   <dbl>
1 TRUE       448     171.    3.78
2 FALSE      818     166.    3.62
```

Among all applicants, under-represented minorities, have lower LSAT scores and GPAs.

```
# A tibble: 2 x 4
   urm2 total avg_lsat avg_gpa
  <dbl> <int>    <dbl>   <dbl>
1     1   212     164.    3.61
2     0  1054     168.    3.69
```

Looking at the cross-tabs, accepted URMs have lower scores/grades than other accepted students.

```
# A tibble: 4 x 5
  accepted  urm2 total avg_lsat avg_gpa
  <lgl>    <dbl> <int>    <dbl>   <dbl>
1 TRUE         1    98     167.    3.70
2 TRUE         0   350     172.    3.80
3 FALSE        1   114     162.    3.52
4 FALSE        0   704     166.    3.63
```

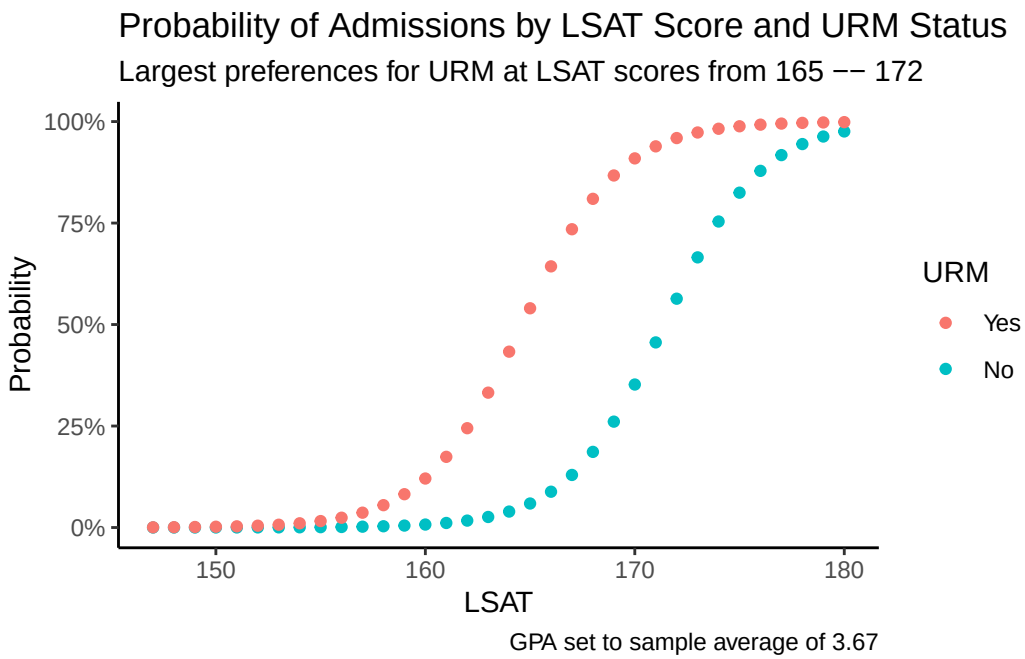
All of this suggests that LSAT, GPA and URM-status plays a role in admissions. A simple Bayesian logistic model confirms that intuition.

```
stan_glm
 family:       binomial [logit]
 formula:      accepted ~ lsat + gpa2 + urm2
 observations: 1266
 predictors:   4
------
            Median MAD_SD
(Intercept) -91.0    5.2
lsat          0.4    0.0
gpa2          4.6    0.4
urm2          2.9    0.3

------
* For help interpreting the printed output see ?print.stanreg
* For info on the priors used see ?prior_summary.stanreg
```

The three predictors all have the expected sign and are all highly statistically significant. Are the coefficients robust to other modelling choices? Tough to say. I found no evidence that interaction terms are important. However, if we defines `accepted` to include anyone accepted (either regular or off the waitlist) or placed on the waitlist at the beginning, we get a different model.

```
stan_glm
 family:       binomial [logit]
 formula:      accepted ~ lsat + gpa2 + urm2
 observations: 1266
 predictors:   4
------
            Median MAD_SD
(Intercept) -42.4    3.1
lsat          0.2    0.0
gpa2          3.1    0.3
urm2          1.5    0.2

------
* For help interpreting the printed output see ?print.stanreg
* For info on the priors used see ?prior_summary.stanreg
```

All the coeffients have the correct sign and are significant. But all are *smaller*. For me, this is a sign that the waitlist decision is much noisier, subject to various weirdnesses. Including these applicants as accepted *dilutes* the signal, as if we just added random noice to the process. So, I will stick with the first model.


Probability of Admissions by LSAT Score and URM Status
Largest preferences for URM at LSAT scores from 165 −− 172
100%
75%
50%
25%
0%
Probability
150
160
170
180
LSAT
URM
Yes
No
GPA set to sample average of 3.67