Bayesian Statistics and Modeling
Part III

Bayesian linear models

I love (for better or worse) rooting for the Seahawks. In many recent years, they have insisted on running a LOT more than most teams.

(Plot is frequency of Run-Run-Pass sequences by team)

TEAM	EPA	SUCCESS	FREQUENCY
Seattle	+0.17	41.2%	26%
Tennessee	-0.23	41.3	24
Buffalo	-0.26	43.9	21
L.A. Chargers	-0.13	41.2	20
San Francisco	-0.37	33.3	20
Houston	-0.32	38.9	18
Miami	-0.50	22.6	18
Denver	-0.47	32.4	17
L.A. Rams	**+0.28**	**60.0**	**16**
N.Y. Giants	+0.23	51.5	16
Indianapolis	-0.03	45.5	16
Minnesota	-0.28	41.9	16
Jacksonville	+0.05	40.0	16
Oakland	-0.72	33.3	16
Cleveland	+0.37	46.7	15
Chicago	-0.09	41.4	15
Pittsburgh	+0.70	61.5	14
Atlanta	+0.37	51.7	14
Detroit	+0.00	50.0	14
Tampa Bay	+0.44	47.8	14
New Orleans	+0.04	41.7	14
Arizona	-0.71	33.3	14
N.Y. Jets	+0.19	50.0	13
Dallas	+0.15	46.4	13
Baltimore	+0.32	44.4	12
Carolina	-0.14	40.9	12
New England	+0.03	39.1	12
Washington	-0.32	34.8	12
Cincinnati	-0.26	47.4	10
Green Bay	-0.10	40.0	10
Kansas City	**+1.19**	**53.3**	**9**
Philadelphia	+0.66	50.0	9

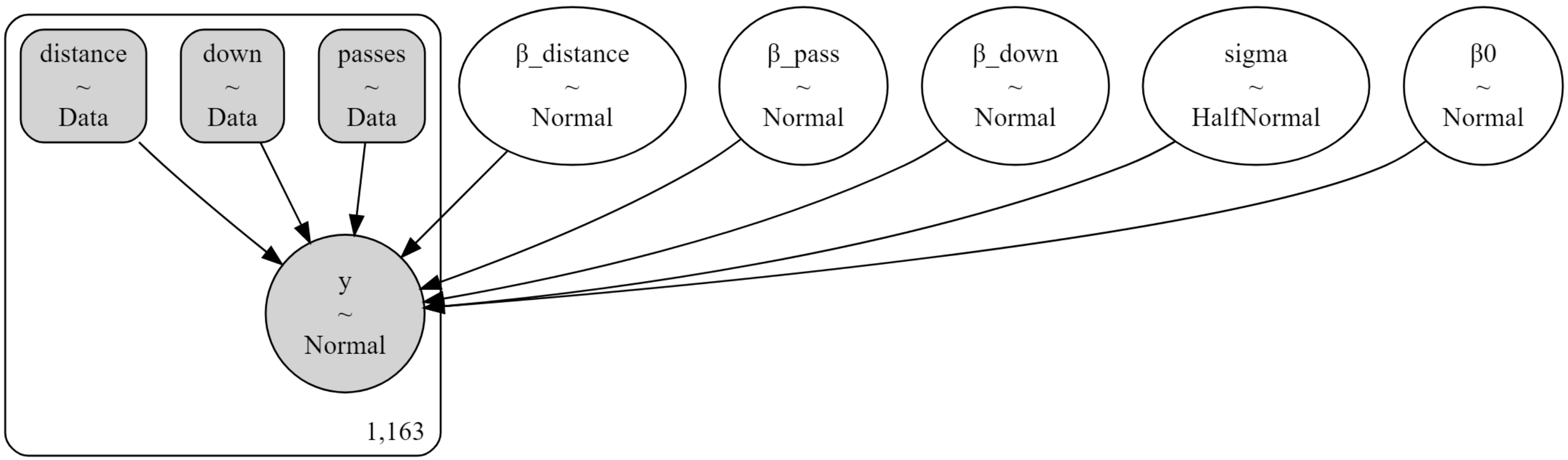
SOURCES: NFL, ELIAS SPORTS BUREAU

Bayesian Linear Models

Can we determine the likelihood of a play being successful based on various characteristics of that play?

- Are runs more successful than passes? (unlikely, but Pete Carroll thinks so)
- We should probably also account for down and distance

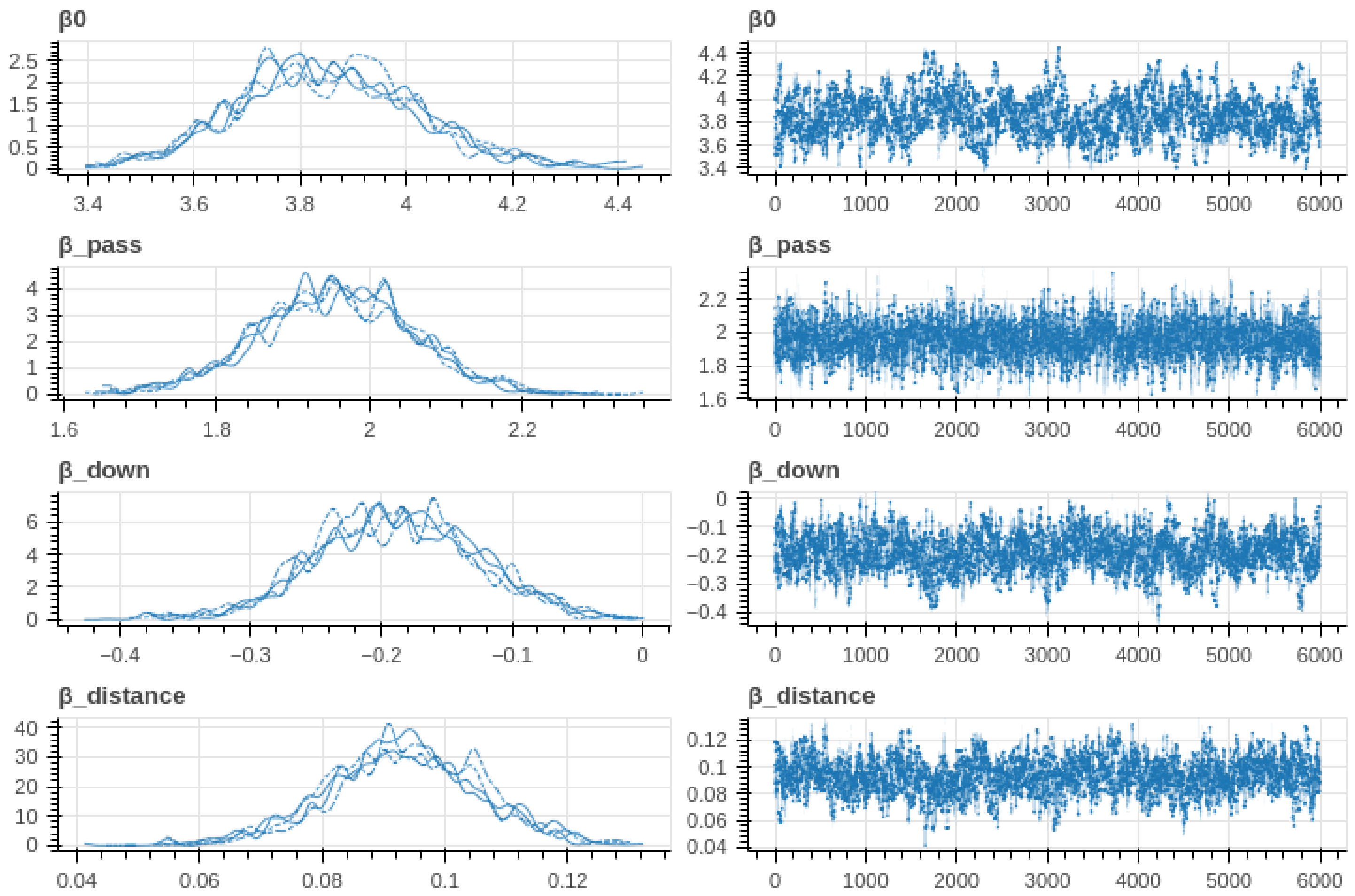
What's the model?



What's the model?

Let's go look at our code now, and generate a regression model using the Bayesian method

Complete Regression Results

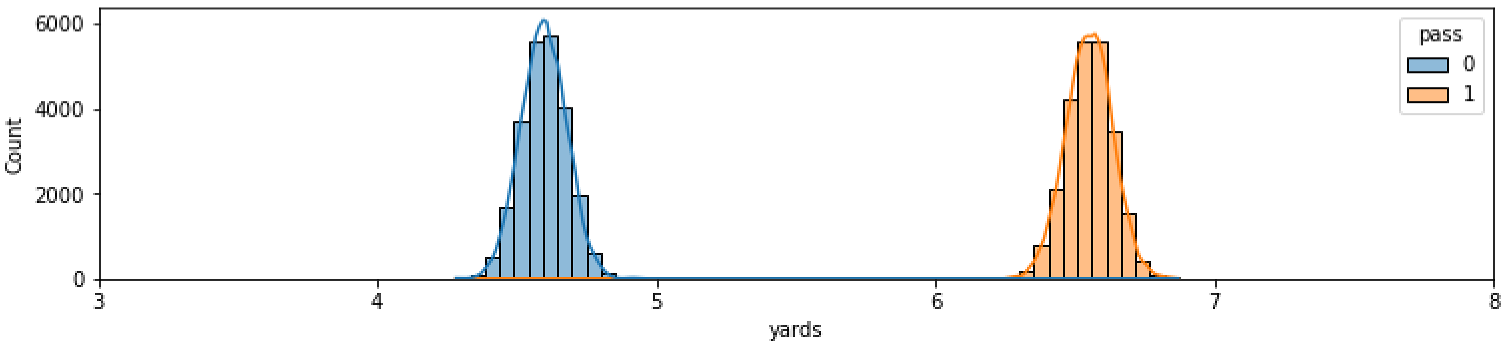


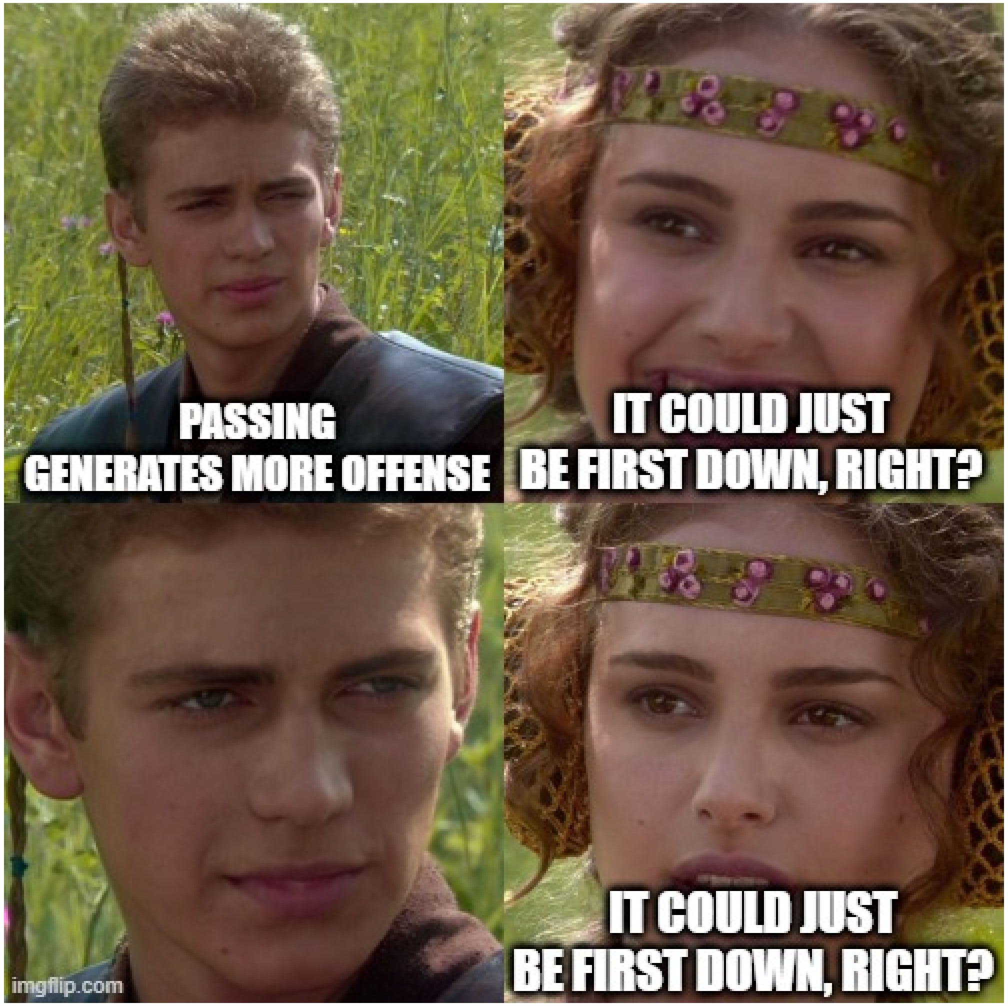
Making sense of choices

What if we want to be able to look at specific contexts?

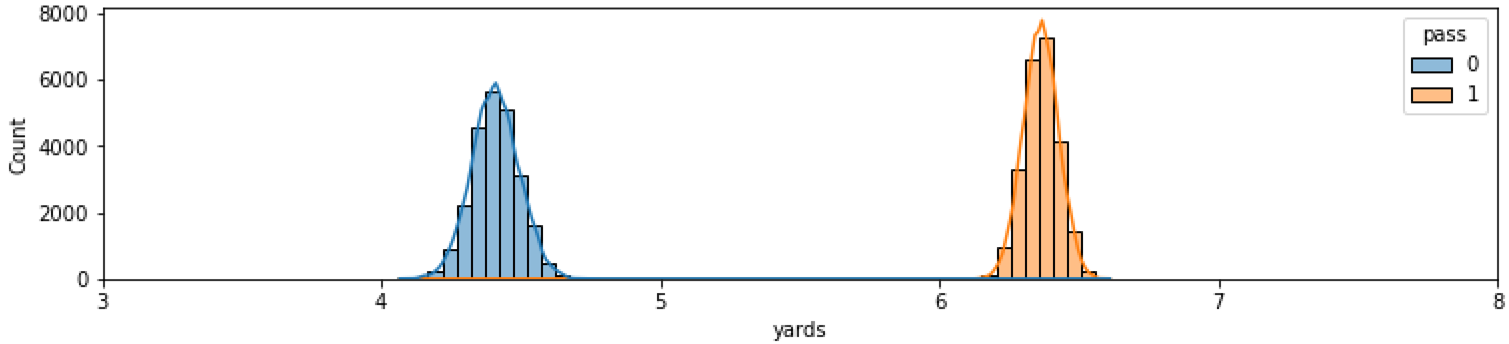
We write quick function and are off to the races!

1st and 10...

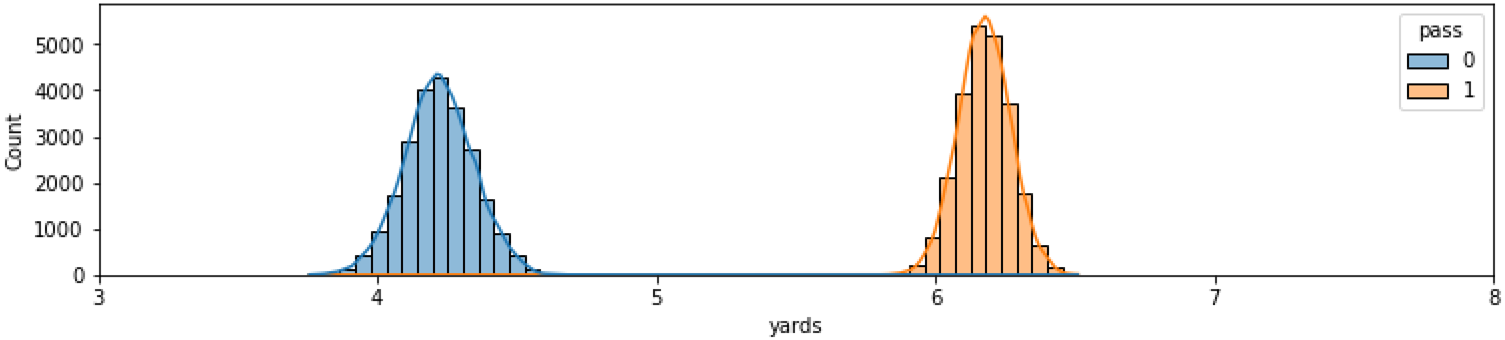



PASSING
GENERATES MORE OFFENSE
IT COULD JUST
BE FIRST DOWN, RIGHT?
imgflip.com
IT COULD JUST
BE FIRST DOWN, RIGHT?

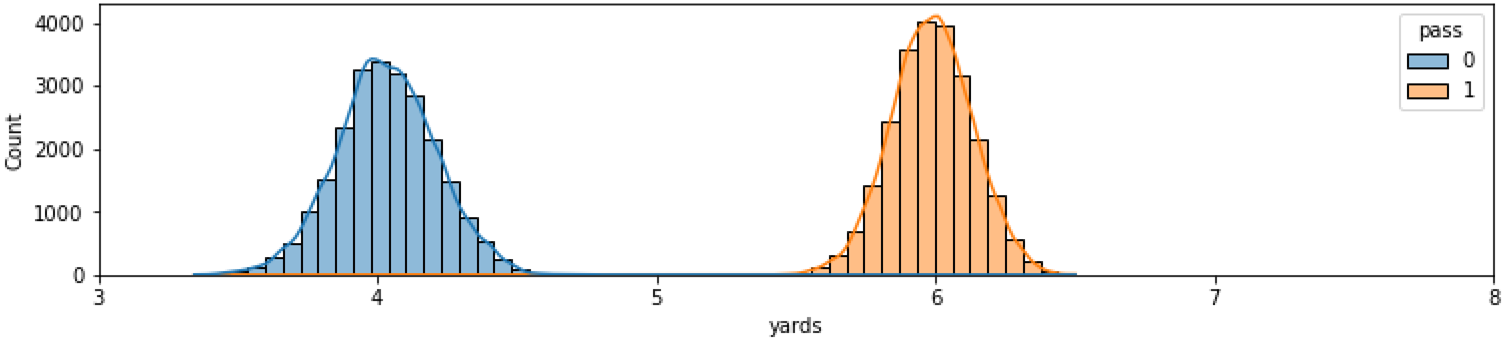
2nd and 10...



3rd and 10...



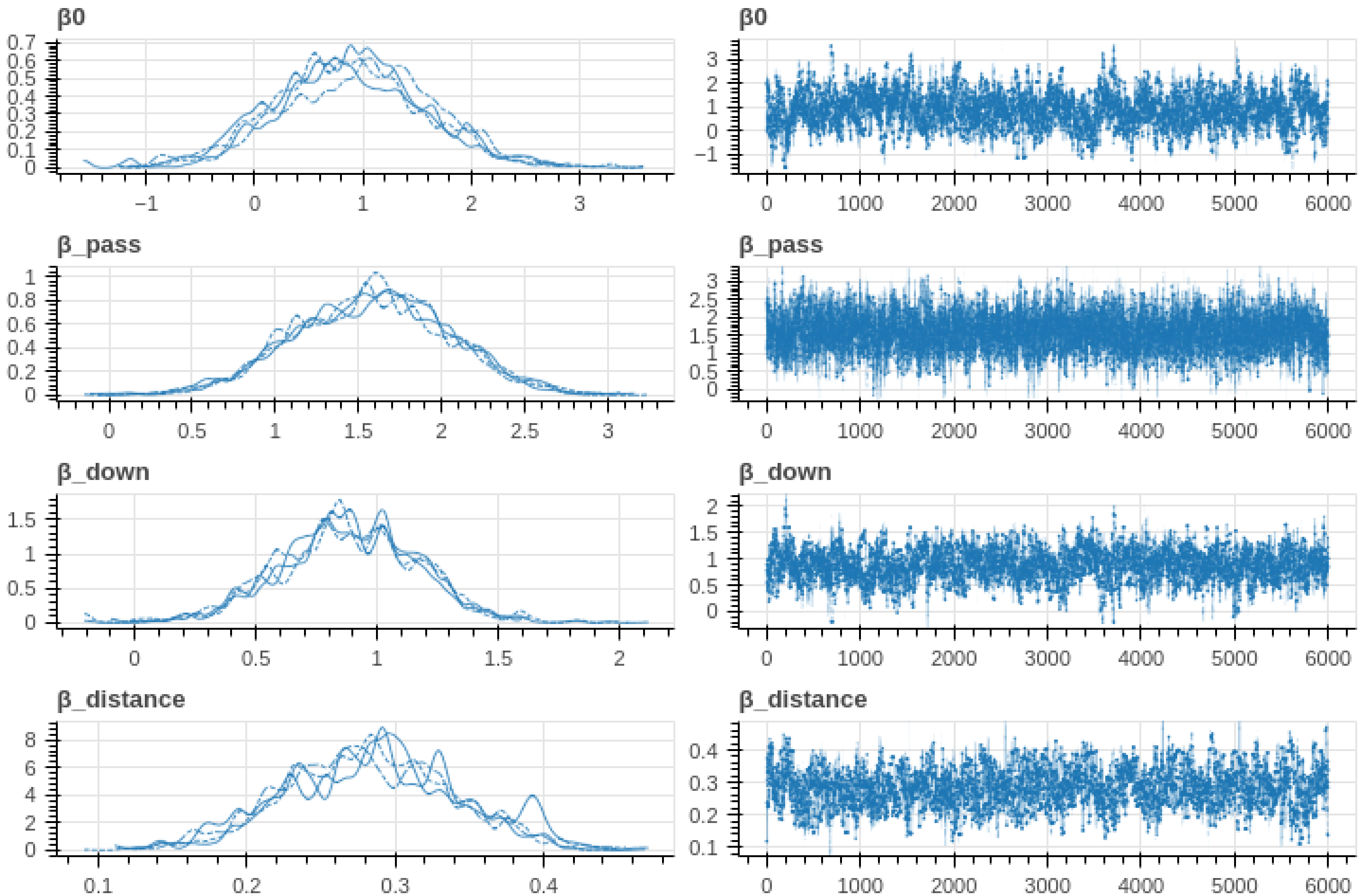
4th and 10!!



Ok, but maybe the Seahawks are different!

Are they? Let's estimate our model with only Seahawks data

Complete Regression Results

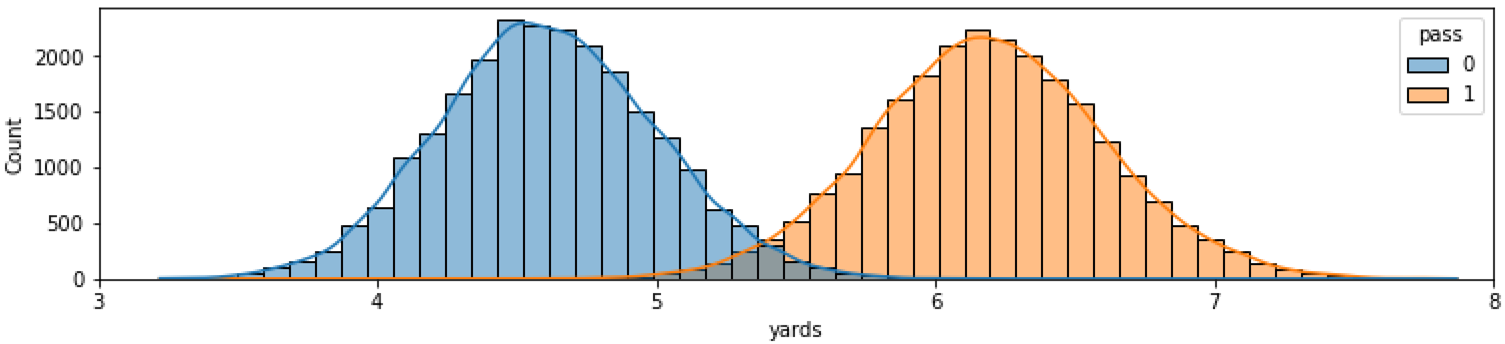


Making sense of choices

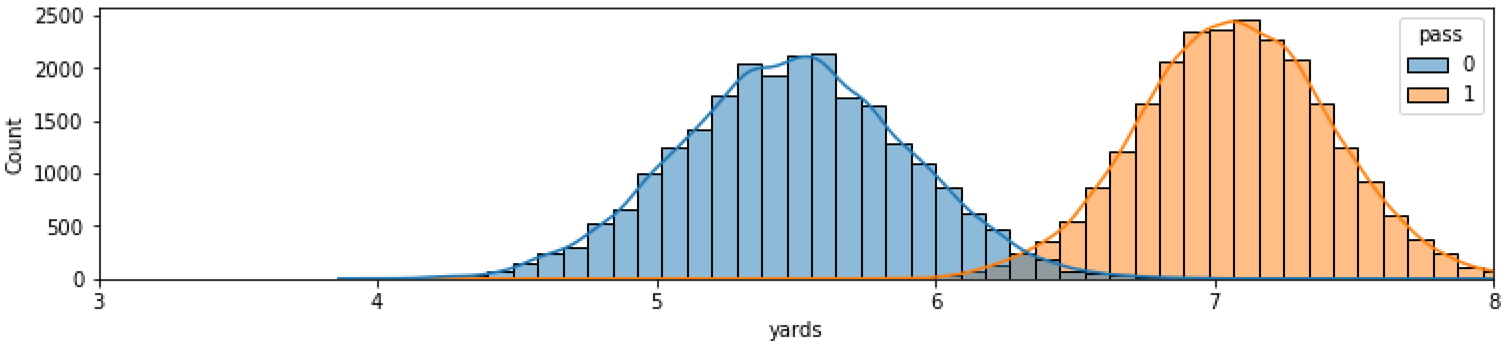
What if we want to be able to look at specific contexts?

We write quick function and are off to the races!

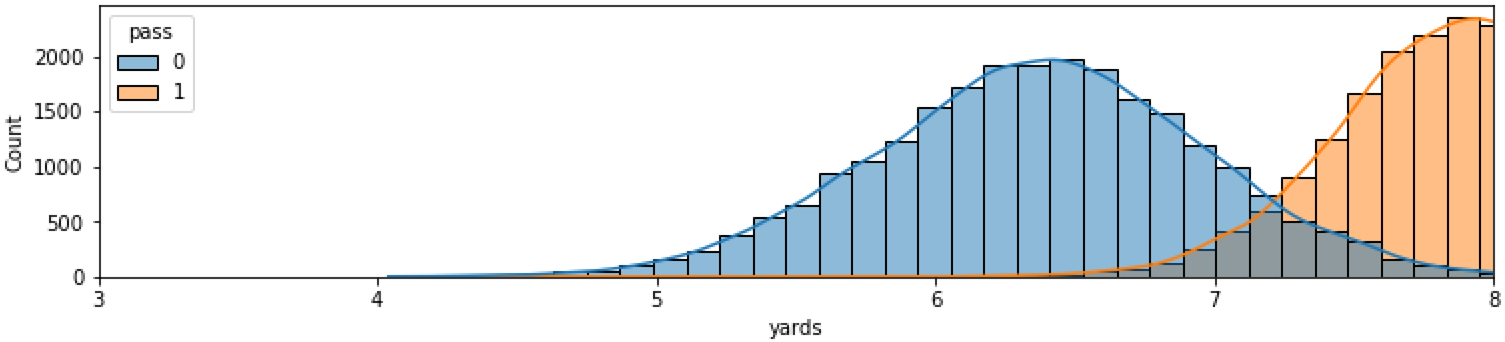
1st and 10...



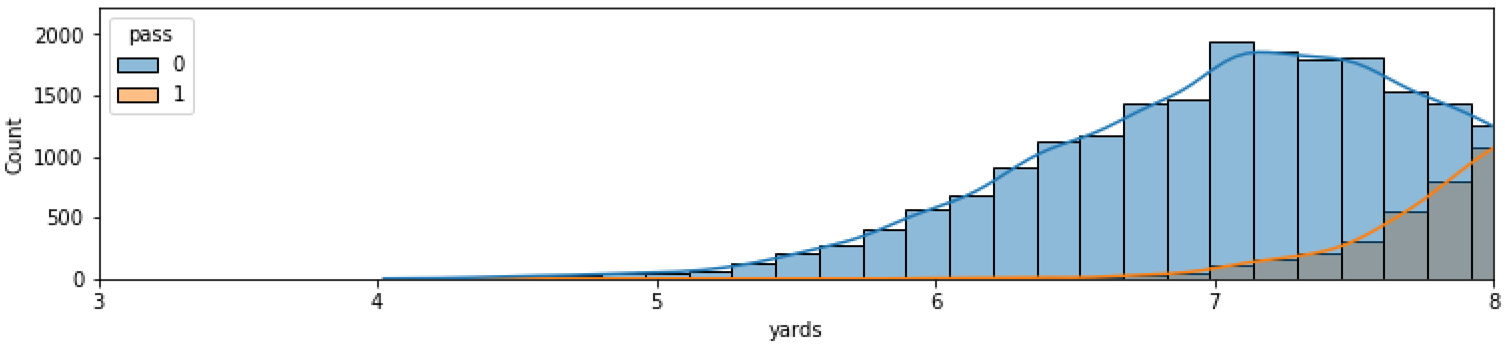
2nd and 10...



3rd and 10...



4th and 10!!



How does it work?

1. Our model creates a `trace` object
2. Each trace contains however many samples (in this case ~40k) of the estimated parameter
3. We use these to look at the **distribution of parameter values**

Credible intervals

Rather than having Confidence Intervals, we have Credible Intervals in Bayesian statistics.

- 95% of sampled parameter values fall inside a 95% CI
- We can shape them arbitrarily
- We can also just use them to measure the likelihood that one measure exceeds another!
    - For example, our distributions for the seahawks overlap, but that doesn't mean running is EVER better!

More flexibility

We have only scratched the surface, but we are starting to see how we can create flexible models that allow us to ask much more **real** questions than we might with a null-hypothesis framework

Lab Time!