Tim Hopp
7/25/2021
IST 718

Lab 1: Calculating Coaches Pay

1. The Data: The final dataframe I used to build my linear model consisted of 123 rows and 15 columns. Notable columns included: TotalPay, Conference, StadiumSize, WinPct and GamesPlayed.

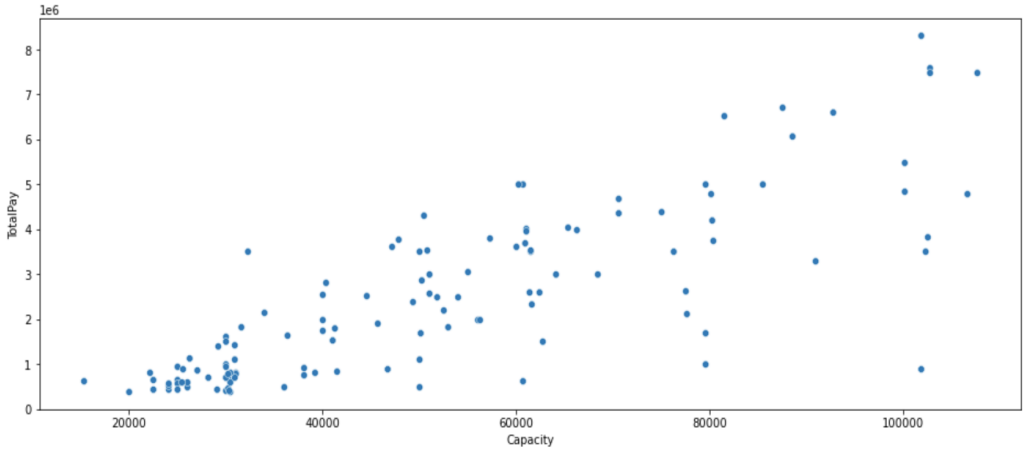
	Coach	TotalPay	Conference	SchoolPay	Bonus	Stadium	Capacity	Won	Lost	Pct.	Total Games	Years	StadiumS
0	Troy Calhoun	885000	Mountain West	885000	247000	Falcon Stadium	46692	404.0	332.0	0.548	749.0	64.0	Medi
1	Terry Bowden	412500	MAC	411000	$225,000	InfoCision Stadium–Summa Field	30000	524.0	563.0	0.483	1123.0	119.0	Sm
2	Nick Saban	8307000	SEC	8307000	$1,100,000	Bryant–Denny Stadium	101821	929.0	331.0	0.729	1303.0	125.0	La
3	Bill Clark	900000	C-USA	900000	$950,000	Bryant–Denny Stadium	101821	606.0	554.0	0.522	1207.0	120.0	La
19	Scott Satterfield	712500	Sun Belt	712500	$295,000	Kidd Brewer Stadium	30000	639.0	339.0	0.649	1007.0	90.0	Sm
20	Kevin Sumlin	2000000	Pac-12	1600000	$2,025,000	Arizona Stadium	56029	617.0	478.0	0.562	1128.0	116.0	Medi
22	Herm Edwards	2000000	Pac-12	2000000	$3,010,000	Frank Kush Field at Sun Devil Stadium	56232	624.0	401.0	0.606	1049.0	107.0	Medi
23	Chad Morris	3500000	SEC	3500000	$1,000,000	Donald W. Reynolds Razorback Stadium, Frank Br...	76212	720.0	521.0	0.578	1281.0	126.0	La
24	Blake Anderson	825000	Sun Belt	825000	$185,000	Centennial Bank Stadium	30964	485.0	499.0	0.493	1021.0	105.0	Medi

The above dataframe will be used to construct a linear model to predict the salary of the next head Football coach at Syracuse University. Below we will start by exploring the variables visually.

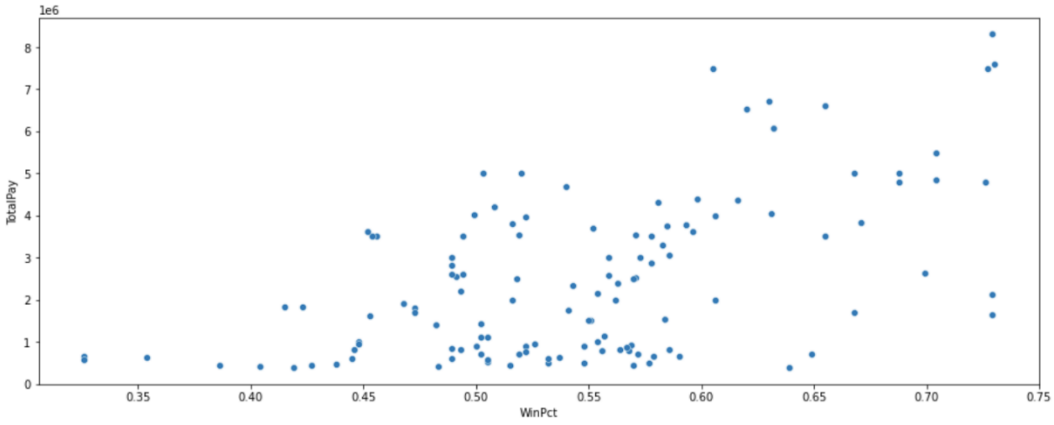
Tim Hopp
7/25/2021
IST 718

2. Exploratory Data Analysis: Before constructing the linear model, we will benefit from looking at the shape/relationships of some of our variables. After creating some initial visualizations, it appears that several variables will be important when constructing the model:

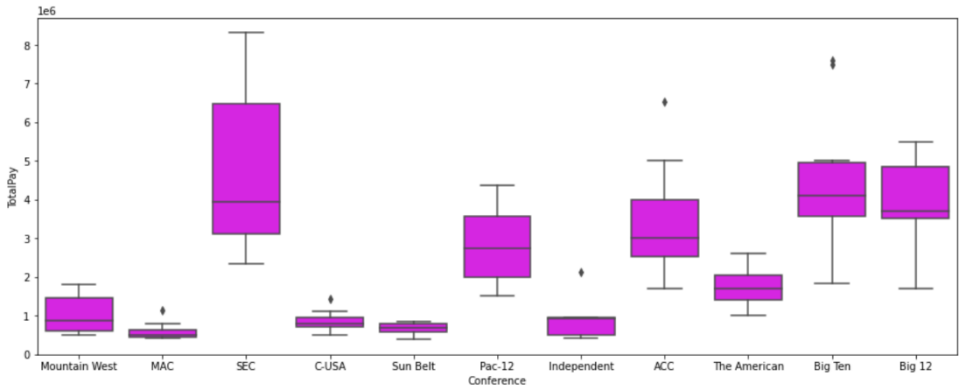
    a. **Stadium Capacity**: we can see below that the larger the stadium capacity the higher the coach is paid.



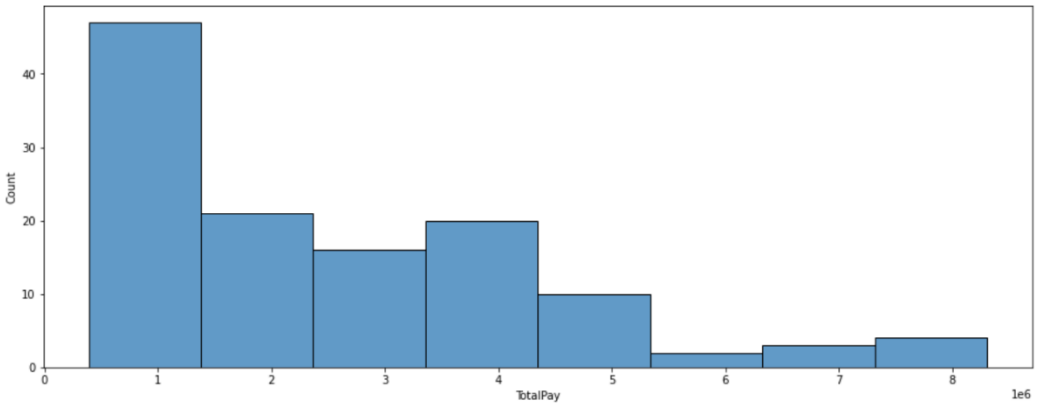
    b. Win Percentage: Logically we would expect that the more wins a coach has, the higher their salary will be. We can see this proven in the plot below:



c. Conference: Below we can see the median salary of head coach per conference. From this visual we can see that there is a high disparity of salary between conference:



d. TotalPay: Lastly, we should look at this distribution of coaches' salaries. It appears that most coaches make ~$1,000,000 and very few coaches make more than $6,000,000:



Tim Hopp
7/25/2021
IST 718

3. Modelling Coaches Salary: I began modelling head coaches' salaries by including all variables and using a step-down regression process after each iteration to include only statistically significant variables.
    a. The initial model (TotalPay ~ Conference + Capacity + Won + Lost + WinPct + Years + GamesPlayed) is summarized below:

OLS Regression Results

Dep. Variable:	TotalPay	R-squared:	0.805
Model:	OLS	Adj. R-squared:	0.776
Method:	Least Squares	F-statistic:	27.38
Date:	Sun, 25 Jul 2021	Prob (F-statistic):	1.98e-30
Time:	10:30:35	Log-Likelihood:	-1851.0
No. Observations:	123	AIC:	3736.
Df Residuals:	106	BIC:	3784.
Df Model:	16		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	3.76e+06	1.47e+06	2.555	0.012	8.42e+05	6.68e+06
Conference[T.Big 12]	4.058e+05	3.84e+05	1.055	0.294	-3.57e+05	1.17e+06
Conference[T.Big Ten]	5.291e+05	3.77e+05	1.404	0.163	-2.18e+05	1.28e+06
Conference[T.C-USA]	-1.867e+06	3.88e+05	-4.805	0.000	-2.64e+06	-1.1e+06
Conference[T.Independent]	-2.048e+06	5.35e+05	-3.825	0.000	-3.11e+06	-9.87e+05
Conference[T.MAC]	-1.978e+06	4.21e+05	-4.695	0.000	-2.81e+06	-1.14e+06
Conference[T.Mountain West]	-1.655e+06	3.76e+05	-4.397	0.000	-2.4e+06	-9.09e+05
Conference[T.Pac-12]	-5.347e+05	3.72e+05	-1.436	0.154	-1.27e+06	2.03e+05
Conference[T.SEC]	7.914e+05	3.61e+05	2.192	0.031	7.57e+04	1.51e+06
Conference[T.Sun Belt]	-1.969e+06	4.45e+05	-4.427	0.000	-2.85e+06	-1.09e+06
Conference[T.The American]	-1.1e+06	4.04e+05	-2.722	0.008	-1.9e+06	-2.99e+05
Capacity	17.4202	6.667	2.613	0.010	4.203	30.637
Won	1.319e+04	1.17e+04	1.124	0.264	-1.01e+04	3.65e+04
Lost	4695.2521	1.15e+04	0.410	0.683	-1.8e+04	2.74e+04
WinPct	-3.544e+06	2.87e+06	-1.236	0.219	-9.23e+06	2.14e+06
Years	1.819e+04	2.24e+04	0.814	0.418	-2.61e+04	6.25e+04
GamesPlayed	-1.053e+04	1.12e+04	-0.943	0.348	-3.26e+04	1.16e+04

Omnibus:	5.411	Durbin-Watson:	1.934
Prob(Omnibus):	0.067	Jarque-Bera (JB):	5.356
Skew:	0.339	Prob(JB):	0.0687
Kurtosis:	3.765	Cond. No.	2.25e+06

    b. After running several iterations of the above regression and removing insignificant variables at every iteration the final regression resulted like so:

OLS Regression Results

Dep. Variable:	TotalPay	R-squared:	0.800
Model:	OLS	Adj. R-squared:	0.776
Method:	Least Squares	F-statistic:	33.49
Date:	Sun, 25 Jul 2021	Prob (F-statistic):	4.34e-32
Time:	11:06:46	Log-Likelihood:	-1852.7
No. Observations:	123	AIC:	3733.
Df Residuals:	109	BIC:	3773.
Df Model:	13		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	1.951e+06	5.71e+05	3.415	0.001	8.19e+05	3.08e+06
Conference[T.Big 12]	4.307e+05	3.82e+05	1.127	0.262	-3.26e+05	1.19e+06
Conference[T.Big Ten]	6.899e+05	3.42e+05	2.019	0.046	1.26e+04	1.37e+06
Conference[T.C-USA]	-1.81e+06	3.82e+05	-4.736	0.000	-2.57e+06	-1.05e+06
Conference[T.Independent]	-1.909e+06	5.15e+05	-3.704	0.000	-2.93e+06	-8.87e+05
Conference[T.MAC]	-1.861e+06	3.79e+05	-4.910	0.000	-2.61e+06	-1.11e+06
Conference[T.Mountain West]	-1.649e+06	3.72e+05	-4.437	0.000	-2.39e+06	-9.13e+05
Conference[T.Pac-12]	-4.516e+05	3.47e+05	-1.302	0.196	-1.14e+06	2.36e+05
Conference[T.SEC]	8.332e+05	3.55e+05	2.348	0.021	1.3e+05	1.54e+06
Conference[T.Sun Belt]	-1.845e+06	4.35e+05	-4.240	0.000	-2.71e+06	-9.82e+05
Conference[T.The American]	-1.07e+06	4.01e+05	-2.671	0.009	-1.86e+06	-2.76e+05
Capacity	17.0408	6.603	2.581	0.011	3.954	30.127
Won	5170.0510	1241.468	4.164	0.000	2709.501	7630.601
GamesPlayed	-2542.8786	721.195	-3.526	0.001	-3972.263	-1113.494

Omnibus:	5.117	Durbin-Watson:	1.937
Prob(Omnibus):	0.077	Jarque-Bera (JB):	5.281
Skew:	0.294	Prob(JB):	0.0713
Kurtosis:	3.828	Cond. No.	6.50e+05

    c. Lastly, we should see if any of the variables are highly correlated and influencing our regression unduly. We can see a very strong positive correlation between GamesPlayed and total games Won at .89 as well as a strong positibe correlation between Capacity and total games Won at .67. Due to the small number of dependent variables being used I will keep these variables despite their high correlation.

Tim Hopp
7/25/2021
IST 718

	TotalPay	SchoolPay	Capacity	Won	Lost	Pct.	Total Games	Years	WinPct	GamesPlayed
TotalPay	1.000000	0.999689	0.808542	0.648167	0.068655	0.547748	0.473676	0.415314	0.547748	0.473676
SchoolPay	0.999689	1.000000	0.808285	0.647948	0.068654	0.547404	0.473517	0.415092	0.547404	0.473517
Capacity	0.808542	0.808285	1.000000	0.665960	0.004241	0.627025	0.453970	0.382675	0.627025	0.453970
Won	0.648167	0.647948	0.665960	1.000000	0.414167	0.707221	0.886818	0.837781	0.707221	0.886818
Lost	0.068655	0.068654	0.004241	0.414167	1.000000	-0.261046	0.787240	0.824794	-0.261046	0.787240
Pct.	0.547748	0.547404	0.627025	0.707221	-0.261046	1.000000	0.344594	0.261080	1.000000	0.344594
Total Games	0.473676	0.473517	0.453970	0.886818	0.787240	0.344594	1.000000	0.986107	0.344594	1.000000
Years	0.415314	0.415092	0.382675	0.837781	0.824794	0.261080	0.986107	1.000000	0.261080	0.986107
WinPct	0.547748	0.547404	0.627025	0.707221	-0.261046	1.000000	0.344594	0.261080	1.000000	0.344594
GamesPlayed	0.473676	0.473517	0.453970	0.886818	0.787240	0.344594	1.000000	0.986107	0.344594	1.000000

Tim Hopp
7/25/2021
IST 718

4. Predicting Syracuse's Head Coach Salary: The final equation used to predict the Syracuse Head Football Coach Salary is:

   $1,951,000 (Intercept/ACC Conference) + $17.04(Capacity) + $5,170(Games Won) + -$2,543(Games Played)

   FINAL SALARY RECOMMENDATION FOR SYRACUSE:

   $1,951,000*(1) + 17.04(49,250) + 5170(725) – 2543(1331) =

   # $1,397,837

5. What would Syracuse pay their head football coach is they were a member of the Big East? The Big Ten?
   a. To determine the answer to this question I calculated the mean salary of all conferences (magenta bar plot on pg. 2) and then divided the ACC mean salary of $3.3 against the specific conferences mean salary to determine a conference coefficient multiplier for salary.
   b. If Syracuse were to join the Big Ten conference we would expect to pay the head football coach: $1,813,119 - a premium of $415,282 compared to the predicted Syracuse salary. This calculation was derived by the following:
      i. (B1G mean salary) / (ACC mean salary) * (Syracuse predicted salary)
         OR
      ii. ($4,304,014 / $3,318,210) * ($1,397,837) = $1,813,119
   c. If Syracuse were to join a much smaller conference like the Sun Belt Conference, we would only expect to pay the head football coach $274,094. This calculation was derived by the following:
      i. (Sun Belt mean salary) / (ACC mean salary) * (Syracuse predicted salary)
         OR
      ii. ($650,650 / $3,318,210) * ($1,397,837) = $274,094

Tim Hopp
7/25/2021
IST 718

6. Further Questions to answer:
    a. What schools did we drop from our data and why? – I did not drop any schools from the data. I chose to not drop any schools from the data because the linear model was built using each schools conference. It did appear that the Big 12 and Pac 12 conferences were statistically not significant meaning that predicting salaries for schools in those conferences could be unreliable.
    b. I did not use Graduation Rate as a variable and used other continuous variables such as Stadium Capacity and # of Games Won. Out of these two outside variables I pulled into the data set provided for class it appears that Wins are much more important for predicting salary than Stadium Capacity with a coefficient of $5,170 vs a coefficient of $17 respectively.
    c. Our final adjusted R^2 is .776 meaning our model accounts for ~78% of the variance in the data. Having an adjusted R^2 this high indicates that our model fits the data quite well.
    d. The single largest impact on coach's salary is determined by the conference they coach in. Large, popular conferences such as the SEC and Big Ten have large stadiums, a storied and long history playing football and an impassioned fan base to attend these games and cheer these teams on. Smaller conferences don't attract as large of fanbases or viewership numbers and typically contain teams that on average are 40 years newer than large schools in large conferences.