Marathon Data Set

By: Ashleigh, Natalie, and Olivia

Objectives

1. Based on our dataset, what variables influence marathon time and what variable has the strongest influence on marathon time?
2. Do injury and footwear interact to affect marathon time?
3. Can we accurately predict gender based on age, marathon time, weekly mileage and peak mileage?

What Influences Marathon Time?

Data Analysis Type: **Multiple Linear Regression**

Variables:

Y = **Marathon Time**

X_1 = **Age**

X_2 = **Gender**

X_3 = **bmi**

X_4 = **Weekly Mileage**

X_5 = **Peak Mileage**

X_6 = **Marathon Fitness**

X_7 = **Footwear**

X_8 = **Injury**

Response Marathon Time

Whole Model

Summary of Fit

RSquare	0.571426
RSquare Adj	0.566018
Root Mean Square Error	29.43156
Mean of Response	224.4658
Observations (or Sum Wgts)	964

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	12	1098351.0	91529.3	105.6655
Error	951	823772.2	866.2	Prob > F
C. Total	963	1922123.2		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	155.24213	10.70111	14.51	<.0001*
age	0.8394545	0.107079	7.84	<.0001*
Gender[0]	-15.89261	1.060096	-14.99	<.0001*
bmi	4.7448012	0.361177	13.14	<.0001*
injury[1]	-1.326279	1.487286	-0.89	0.3728
injury[2]	-3.519069	1.687323	-2.09	0.0373*
footwear[1]	-10.32045	3.00151	-3.44	0.0006*
footwear[2]	-6.262343	2.815066	-2.22	0.0263*
Marathon Fitness[1]	-11.07294	2.54633	-4.35	<.0001*
Marathon Fitness[2]	-6.064288	2.396364	-2.53	0.0115*
Marathon Fitness[3]	0.9823164	3.869712	0.25	0.7997
Weekly Mileage	-1.020893	0.128241	-7.96	<.0001*
Peak Mileage	-0.199946	0.105888	-1.89	0.0593

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
age	1	1	53236.83	61.4590	<.0001*
Gender	1	1	194682.54	224.7504	<.0001*
bmi	1	1	149493.59	172.5822	<.0001*
injury	2	2	4401.41	2.5406	0.0794
footwear	2	2	10256.34	5.9202	0.0028*
Marathon Fitness	3	3	16616.12	6.3941	0.0003*
Weekly Mileage	1	1	54895.45	63.3738	<.0001*
Peak Mileage	1	1	3088.56	3.5656	0.0593

Residual by Predicted Plot

Summary of Fit

RSquare	0.356178
RSquare Adj	0.354166
Root Mean Square Error	35.90359
Mean of Response	224.4658
Observations (or Sum Wgts)	964

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	684618.0	228206	177.0318
Error	960	1237505.2	1289	**Prob > F**
C. Total	963	1922123.2		<.0001*

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	930	1204037.7	1294.66	1.1605
Pure Error	30	33467.5	1115.58	**Prob > F**
Total Error	960	1237505.2		0.3187
				Max RSq
				0.9826

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	27.105177	10.43579	2.60	0.0095*
age	0.8319484	0.12952	6.42	<.0001*
bmi	7.3629888	0.412921	17.83	<.0001*
Gender[0]	-20.50976	1.245513	-16.47	<.0001*

0.9826

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	27.813743	11.0019	2.53	0.0116*
age	0.7964404	0.146024	5.45	<.0001*
bmi	7.3648863	0.416941	17.66	<.0001*
Gender[0]	-20.44938	1.280023	-15.98	<.0001*
(age-37.3071)*(bmi-23.3692)	0.0610712	0.049804	1.23	0.2204
(bmi-23.3692)*Gender[0]	0.5741155	0.423511	1.36	0.1755
(age-37.3071)*Gender[0]	0.0468547	0.149637	0.31	0.7543

Effect Tests

Response Marathon Time

Whole Model

Actual by Predicted Plot

Summary of Fit

RSquare	0.569819
RSquare Adj	0.564848
Root Mean Square Error	29.47119
Mean of Response	224.4658
Observations (or Sum Wgts)	964

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	11	1095262.5	99569.3	114.6384
Error	952	826860.7	868.6	**Prob > F**
C. Total	963	1922123.2		<.0001*

Parameter Estimates

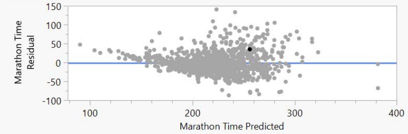
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	152.57928	10.62207	14.36	<.0001*
age	0.8480613	0.107126	7.92	<.0001*
Gender[0]	-16.05314	1.058105	-15.17	<.0001*
bmi	4.7653897	0.361498	13.18	<.0001*
injury[1]	-1.37053	1.489104	-0.92	0.3576
injury[2]	-3.559294	1.689461	-2.11	0.0354*
footwear[1]	-10.66026	3.000144	-3.55	0.0004*
footwear[2]	-6.280326	2.818841	-2.23	0.0261*
Marathon Fitness[1]	-11.01254	2.549558	-4.32	<.0001*
Marathon Fitness[2]	-5.964113	2.399003	-2.49	0.0131*
Marathon Fitness[3]	0.3151462	3.858737	0.08	0.9349
Weekly Mileage	-1.228299	0.066279	-18.53	<.0001*

Effect Tests

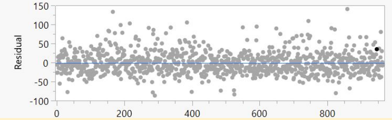
Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
age	1	1	54432.72	62.6707	<.0001*
Gender	1	1	199920.82	230.1774	<.0001*
bmi	1	1	150931.31	173.7736	<.0001*
injury	2	2	4533.67	2.6099	0.0741
footwear	2	2	11015.44	6.3413	0.0018*
Marathon Fitness	3	3	16566.79	6.3580	0.0003*
Weekly Mileage	1	1	298294.35	343.4390	<.0001*

Residual by Predicted Plot

Residual by Predicted Plot



Residual by Row Plot



Marathon Time Predicted

Prediction Expression

152.57928381

+ 0.8480613155 • age

+ Match(Gender) "0" ⇒ -16.05313773; "1" ⇒ 16.053137727; else ⇒ .

+ 4.7653897376 • bmi

+ Match(injury) "1" ⇒ -1.37052964; "2" ⇒ -3.559293751; "3" ⇒ 4.9298233912; else ⇒ .

+ Match(footwear) "1" ⇒ -10.66026162; "2" ⇒ -6.280326091; "3" ⇒ 16.940587707; else ⇒ .

+ Match(Marathon Fitness) "1" ⇒ -11.01253597; "2" ⇒ -5.964112761; "3" ⇒ 0.315146156; "4" ⇒ 16.661502575; else ⇒ .

+ -1.228298574 • Weekly Mileage

Do injury and footwear interact to affect marathon time?

Data Analysis Type: **Two-Way ANOVA with replicates**

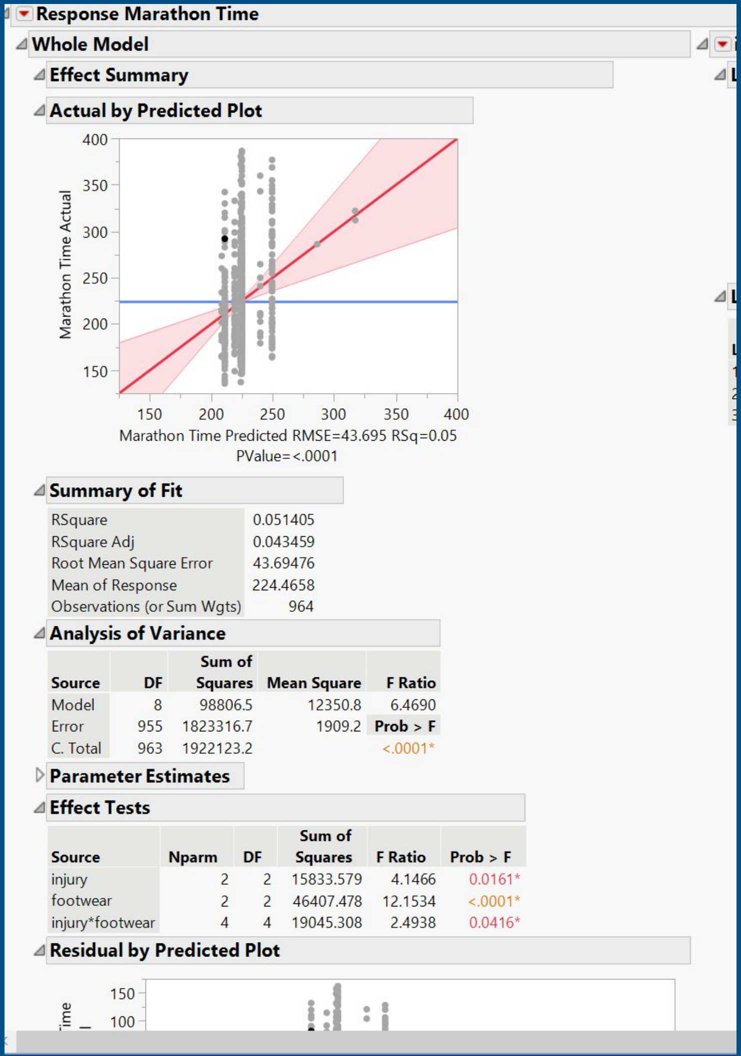
Variables:

Y= **Marathon Time**

X_1= **Injury**

X_2= **Footwear**

X_3= **Injury*Footwear**



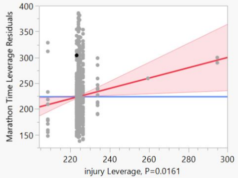
RSquare	0.051405
RSquare Adj	0.043459
Root Mean Square Error	43.69476
Mean of Response	224.4658
Observations (or Sum Wgts)	964

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	8	98806.5	12350.8	6.4690
Error	955	1823316.7	1909.2	Prob > F
C. Total	963	1922123.2		<.0001*

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
injury	2	2	15833.579	4.1466	0.0161*
footwear	2	2	46407.478	12.1534	<.0001*
injury*footwear	4	4	19045.308	2.4938	0.0416*

injury

Leverage Plot

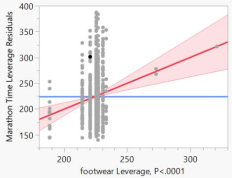


Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
1	225.287	4.61	222.16
2	243.106	14.74	223.24
3	258.377	11.10	244.12

footwear

Leverage Plot

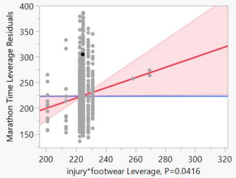


Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
1	212.908	4.44	212.95
2	232.825	2.12	227.08
3	281.037	18.37	254.21

injury*footwear

Leverage Plot



Least Squares Means Table

Level	Least Sq Mean	Std Error
1,1	211.048	3.73
1,2	224.935	2.00
1,3	239.877	13.17
2,1	219.092	6.00
2,2	224.042	3.14

footwear*injury

LSMeans Differences Tukey HSD

LSMean[j]

LSMean[i] / Mean[i]-Mean[j], Std Err Dif, Lower CL Dif, Upper CL Dif	1,1	1,2	1,3	2,1	2,2	2,3	3,1	3,2	3,3
	-13.9299	0	-29.2187	-25.5068	-26.0044	-54.993	-65.793	-204.209	-195.808
	30.01868	0	50.23918	13.82248	16.10473	-5.82114	24.2234	70.02732	-0.10671
1,3	-2.46583	-10.5102	0	-16.3524	-15.46	-40.9173	-31.2951	-77.6011	-108.468
	11.88352	12.7791	0	11.45723	11.70998	12.40158	17.34499	45.12776	32.89222
	-39.4105	-50.2392	0	-51.9718	-51.8652	-79.4726	-85.2189	-217.899	-210.726
	34.47886	29.21872	0	19.26697	20.9451	-2.36203	22.62879	62.69658	-6.20914
2,1	13.88657	5.842168	16.3524	0	0.892353	-24.5649	-14.9427	-61.2487	-92.1154
	4.233423	6.325276	11.45723	0	3.718504	5.522951	13.32488	43.74035	30.9613
	0.725277	-13.8225	-19.267	0	-10.6681	-41.7352	-56.3684	-197.233	-188.371
	27.04786	25.50682	51.97177	0	12.45281	-7.39461	26.48307	74.73566	4.140205
2,2	12.99421	4.949815	15.46005	-0.89235	0	-25.4573	-15.835	-62.1411	-93.0077
	4.876208	6.772345	11.70998	3.718504	0	6.029802	13.54282	43.80723	31.05572
	-2.16543	-16.1047	-20.9451	-12.4528	0	-44.2033	-57.9383	-198.333	-189.557
	28.15386	26.00436	51.86519	10.66811	0	-6.71122	26.26827	74.05125	3.541389
2,3	38.45148	30.40708	40.91731	24.56492	25.45727	0	9.622264	-36.6838	-67.5505
	6.360276	7.908246	12.40158	5.522951	6.029802	0	14.14509	43.99715	31.32305
	18.67802	5.821135	2.362031	7.394614	6.711215	0	-34.3534	-173.467	-164.931
	58.22495	54.99303	79.4726	41.73522	44.20332	0	53.59795	100.099	29.82976
3,1	28.82922	20.78482	31.29505	14.94265	15.835	-9.62226	0	-46.3061	-77.1727
	13.69316	14.47722	17.34499	13.32488	13.54282	14.14509	0	45.63769	33.58843
	-13.7414	-24.2234	-22.6288	-26.4831	-26.2683	-53.5979	0	-188.189	-181.596
	71.39988	65.79304	85.2189	56.36837	57.93828	34.35342	0	95.57696	27.25035
3,2	75.13528	67.09088	77.60111	61.24871	62.14107	36.6838	46.30606	0	-30.8667
	43.85394	44.10505	45.12776	43.74035	43.80723	43.99715	45.63769	0	53.51493
	-61.2022	-70.0273	-62.6966	-74.7357	-74.0512	-100.099	-95.577	0	-197.239
	211.4728	204.2091	217.8988	197.2331	198.3334	173.4665	188.1891	0	135.5059
3,3	106.0019	97.95755	108.4678	92.11538	93.00773	67.55046	77.17273	30.86667	0
	31.12157	31.47442	32.89222	30.9613	31.05572	31.32305	33.58843	53.51493	0
	9.248093	0.106706	6.209145	-4.1402	-3.54139	-29.8298	-27.2503	-135.506	0
	202.7558	195.8084	210.7264	188.371	189.5569	164.9307	181.5958	197.2393	0

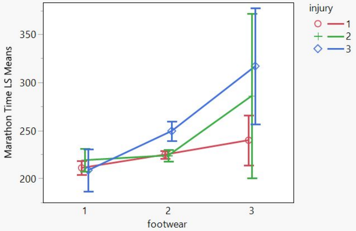
Level					Least Sq Mean	Std Error
3,3	A	B			317.05	30.897
3,2	A	B	C	D	286.18	43.695
2,3	A				249.50	5.149
3,1	A	B	C	D	239.88	13.174
2,1		B	C		224.93	1.996
2,2		B	C	D	224.04	3.137
1,2			C	D	219.09	6.002
1,1				D	211.05	3.733
1,3			C	D	208.58	11.282

Levels not connected by same letter are significantly different.

Level					Least Sq Mean	Std Error
3,3	A	B			317.05	30.897
2,3	A	B	C	D	286.18	43.695
3,2	A				249.50	5.149
1,3	A	B	C	D	239.88	13.174
1,2		B	C		224.93	1.996
2,2		B	C	D	224.04	3.137
2,1			C	D	219.09	6.002
1,1				D	211.05	3.733
3,1			C	D	208.58	11.282

Levels not connected by same letter are significantly different.

Least Squares Means Plot



Prediction Expression

$$
\begin{aligned}
&242.25664005\\
&+\mathrm{Match}(\mathrm{injury})\begin{pmatrix}\text{"1"} \Rightarrow -16.96999105\\ \text{"2"} \Rightarrow 0.8493780113\\ \text{"3"} \Rightarrow 16.120613041\\ \text{else} \Rightarrow .\end{pmatrix}\\
&+\mathrm{Match}(\mathrm{footwear})\begin{pmatrix}\text{"1"} \Rightarrow -29.34906385\\ \text{"2"} \Rightarrow -9.431164776\\ \text{"3"} \Rightarrow 38.780228625\\ \text{else} \Rightarrow .\end{pmatrix}\\
&+\mathrm{Match}(\mathrm{injury})\begin{pmatrix}\text{"1"} \Rightarrow \mathrm{Match}(\mathrm{footwear})\begin{pmatrix}\text{"1"} \Rightarrow 15.110468384\\ \text{"2"} \Rightarrow 9.0791365184\\ \text{"3"} \Rightarrow -24.1896049\\ \text{else} \Rightarrow .\end{pmatrix}\\ \text{"2"} \Rightarrow \mathrm{Match}(\mathrm{footwear})\begin{pmatrix}\text{"1"} \Rightarrow 5.3354986234\\ \text{"2"} \Rightarrow -9.632585239\\ \text{"3"} \Rightarrow 4.297086616\\ \text{else} \Rightarrow .\end{pmatrix}\\ \text{"3"} \Rightarrow \mathrm{Match}(\mathrm{footwear})\begin{pmatrix}\text{"1"} \Rightarrow -20.44596701\\ \text{"2"} \Rightarrow 0.5534487209\\ \text{"3"} \Rightarrow 19.892518286\\ \text{else} \Rightarrow .\end{pmatrix}\\ \text{else} \Rightarrow .\end{pmatrix}
\end{aligned}
$$

Can we predict gender based on age, marathon time, weekly mileage, and peak mileage?

Data Analysis Type: **Logistic Regression**

Variables:

Y = **Gender (Male/Female)**

X_1 = **Marathon Time**

X_2 = **Age**

X_3 = **Weekly Mileage**

X_4 = **Peak Mileage**

Nominal Logistic Fit for Gender

Effect Summary

Source	Logworth	PValue
Marathon Time	24.665	0.00000
age	15.552	0.00000
Weekly Mileage	3.182	0.00066
Peak Mileage	1.289	0.05143

Remove Add Edit Exclude FDR

Converged in Gradient, 4 iterations

Iterations

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	86.69258	4	173.3852	<.0001*
Full	543.17954			
Reduced	629.87212			

RSquare (U)	0.1376
AICc	1096.42
BIC	1120.71
Observations (or Sum Wgts)	964

Fit Details

Lack Of Fit

Source	DF	-LogLikelihood	ChiSquare
Lack Of Fit	959	543.17954	1086.359
Saturated	963	0.00000	Prob>ChiSq
Fitted	4	543.17954	0.0025*

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	-3.7953792	0.7268601	27.27	<.0001*
Marathon Time	0.022789	0.0024238	88.40	<.0001*
Peak Mileage	-0.0166229	0.0085687	3.76	0.0524
age	-0.0695816	0.009103	58.43	<.0001*
Weekly Mileage	0.03594794	0.0106729	11.34	0.0008*

For log odds of 1/0

Covariance of Estimates

Effect Likelihood Ratio Tests

Source	Nparm	DF	L-R ChiSquare	Prob>ChiSq
Marathon Time	1	1	108.430466	<.0001*
Peak Mileage	1	1	3.79414032	0.0514
age	1	1	66.9348368	<.0001*
Weekly Mileage	1	1	11.6038626	0.0007*

Nominal Logistic Fit for Gender

Effect Summary

Source	Logworth	PValue
Marathon Time	25.604	0.00000
age	15.339	0.00000
Weekly Mileage	2.786	0.00164

Remove Add Edit Exclude FDR

Converged in Gradient, 4 iterations

Iterations

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	84.79551	3	169.591	<.0001*
Full	545.07662			
Reduced	629.87212			

RSquare (U)	0.1346
AICc	1098.19
BIC	1117.64
Observations (or Sum Wgts)	964

Fit Details

Lack Of Fit

Source	DF	-LogLikelihood	ChiSquare
Lack Of Fit	958	543.69032	1087.381
Saturated	961	1.38629	Prob>ChiSq
Fitted	3	545.07662	0.0022*

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	-4.057488	0.7128336	32.40	<.0001*
Marathon Time	0.0230728	0.0024129	91.44	<.0001*
age	-0.0688296	0.0090587	57.73	<.0001*
Weekly Mileage	0.01900724	0.0060235	9.96	0.0016*

For log odds of 1/0

Covariance of Estimates

Effect Likelihood Ratio Tests

Source	Nparm	DF	L-R ChiSquare	Prob>ChiSq
Marathon Time	1	1	112.718768	<.0001*
age	1	1	65.9703822	<.0001*
Weekly Mileage	1	1	9.91612519	0.0016*

Final Model

P = **probability of female**

$$P = \frac{e^{-4.057 + 0.023x1 - 0.069x2 + 0.019x3}}{1 + e^{-4.057 + 0.023x1 - 0.069x2 + 0.019x3}}$$

ROC table:

If p >= 0.3395

Predict Female

If p < 0.3395

Predict Male

Nominal Logistic Fit for Gender

Effect Summary

Source	Logworth	PValue
Marathon Time	25.604	0.00000
age	15.339	0.00000
Weekly Mileage	2.786	0.00164

Remove Add Edit Exclude FDR

Converged in Gradient, 4 iterations

Iterations

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	84.79551	3	169.591	<.0001*
Full	545.07662			
Reduced	629.87212			

RSquare (U)	0.1346
AICc	1098.19
BIC	1117.64
Observations (or Sum Wgts)	964

Fit Details

Lack Of Fit

Source	DF	-LogLikelihood	ChiSquare
Lack Of Fit	958	543.69032	1087.381
Saturated	961	1.38629	Prob>ChiSq
Fitted	3	545.07662	0.0022*

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	-4.057488	0.7128336	32.40	<.0001*
Marathon Time	0.0230728	0.0024129	91.44	<.0001*
age	-0.0688296	0.0090587	57.73	<.0001*
Weekly Mileage	0.01900724	0.0060235	9.96	0.0016*

For log odds of 1/0

Covariance of Estimates

Effect Likelihood Ratio Tests

Source	Nparm	DF	L-R ChiSquare	Prob>ChiSq
Marathon Time	1	1	112.718768	<.0001*
age	1	1	65.9703822	<.0001*
Weekly Mileage	1	1	9.91612519	0.0016*

ROC Table

Prob	1-Specificity	Sensitivity	Sens-(1-Spec)	
0.3404	0.3112	0.7233	0.4122	
0.3401	0.3112	0.7262	0.4150	
0.3398	0.3112	0.7291	0.4179	
0.3397	0.3128	0.7291	0.4163	
0.3395	0.3128	0.7320	0.4192	*
0.3393	0.3144	0.7320	0.4176	
0.3391	0.3160	0.7320	0.4159	
0.3391	0.3177	0.7320	0.4143	

Confusion Matrix + Odds Ratio

Model Accuracy:

Sensitivity: Model's ability to predict Female

136/347 = 0.392

Specificity: Model's ability to predict Male

548/617 = 0.888

Interpret Odds Ratio:

OR Marathon Time: For each additional minute a participant runs, the odds the participant is female increase by 2.33%.

OR Age: For each additional year older a participant is, the odds the participant is female decrease by 6.65%.

OR Weekly: For each additional mile a participant adds to their weekly mileage, the odds the participant is female increase by 1.92%

Confusion Matrix

Training

Actual Gender	Predicted Count 1	Predicted Count 0
1	136	211
0	69	548

Actual Gender	Predicted Rate 1	Predicted Rate 0
1	0.392	0.608
0	0.112	0.888

Odds Ratios

For Gender odds of 1 versus 0

Unit Odds Ratios

Per unit change in regressor

Term	Odds Ratio	Lower 95%	Upper 95%	Reciprocal
Marathon Time	1.023341	1.018627	1.028317	0.9771913
age	0.933486	0.916722	0.94989	1.0712536
Weekly Mileage	1.019189	1.007222	1.031325	0.9811723

Questions?