

Listing 1: Linear Regression Output

Call:

```
lm(formula = logwage ~ hgc + college + tenure + age + married,
    data = wages)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.80084	-0.23093	0.02974	0.24627	0.86314

Coefficients:

	Estimate	Std. Error	t value
(Intercept)	0.6385419	0.1458085	4.379
hgc	0.0618155	0.0054274	11.389
collegenot college grad	0.1464113	0.0347664	4.211
tenure	0.0233959	0.0016747	13.970
age	-0.0006992	0.0027606	-0.253
marriedsingle	-0.0238215	0.0178513	-1.334

Pr(>|t|)

(Intercept)	1.26e-05 ***
hgc	< 2e-16 ***
collegenot college grad	2.68e-05 ***
tenure	< 2e-16 ***
age	0.800
marriedsingle	0.182

Signif. codes:

0	***	0.001	**	0.01	*	0.05	.	0.1	1
---	-----	-------	----	------	---	------	---	-----	---

Residual standard error: 0.3465 on 1663 degrees of freedom
(560 observations deleted due to missingness)

Multiple R-squared: 0.1949, Adjusted R-squared: 0.1925

F-statistic: 80.51 on 5 and 1663 DF, p-value: < 2.2e-16

Question 6: About 24.8 percent of the values in "logwage" are missing. This missing values are likely missing completely at random (MCAR), meaning that the missingness is unrelated of any unobserved data..

Question 8:

I have decided to use financial and stock market data to gain a better understanding of potential investments. The dataset I am using includes all stock tickers from the NASDAQ and NYSE, totaling around 5,600 stock tickers.

	wages1	wages2	wage3	wage4
(Intercept)	0.534	0.708	0.534	0.540
	(0.146)	(0.116)	(0.146)	(0.154)
hgc	0.062	0.050	0.062	0.063
	(0.005)	(0.004)	(0.005)	(0.005)
collegenot college grad	0.145	0.168	0.145	0.141
	(0.034)	(0.026)	(0.034)	(0.030)
tenure	0.050	0.038	0.050	0.049
	(0.005)	(0.004)	(0.005)	(0.004)
l(tenure^2)	-0.002	-0.001	-0.002	-0.002
	(0.000)	(0.000)	(0.000)	(0.000)
age	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)

Figure 1: Enter Caption