Problem-1:

Type of Projection	X2	X3
Hard Hat	**1**	**0**
Bump Cap	**0**	**1**
None	**0**	**0**

X_1 is an index reflecting both the weight of the object and the distance it fell
X2 and X3 are indicator variables for the nature of head protection worn at the time of the accident.
The response function is $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$

a) Develop the regression function for each type of protection category
The response function for hard hat is
$=\beta_0 + \beta_1 x_1 + \beta_2(1) + \beta_3(0)$
$=\beta_0 + \beta_1 x_1 + \beta_2$
$E(\text{Hard Hat})=(\beta_0 + \beta_2)+ \beta_1 x_1$

The response function for Bump Cap is
$=\beta_0 + \beta_1 x_1 + \beta_2(0) + \beta_3(1)$
$=\beta_0 + \beta_1 x_1 + \beta_3$
$=(\beta_0 + \beta_3) + \beta_1 x_1$
$E(\text{Bump Cap})=(\beta_0 + \beta_3) + \beta_1 x_1$

The response function for None is
$=\beta_0 + \beta_1 x_1 + \beta_2(0) + \beta_3(0)$
$=\beta_0 + \beta_1 x_1$
$E(\text{None})=\beta_0 + \beta_1 x_1$

b) For each of the following questions, specify the H_0 and H_a for the appropriate test

1. Null hypothesis: $\beta_3 = 0$ and Alternate Hypothesis: $\beta_3 \neq 0$
2. Null hypothesis: $\beta_3 = \beta_2$ and Alternate Hypothesis: $\beta_3 \neq \beta_2$

Problem-2:

Consider the automobile gasoline mileage data in Table B.3 .
a. Build a linear regression model relating gasoline mileage y to vehicle weight x10 and the type of transmission x11. Does the type of transmission significantly affect the mileage performance?

Solution:

Here,

We are building the linear regression model relating to gasoline mileage y to vehicle weight X10 and type of transmission X11

here we are considering the type of transmission for X11 =1 if A and 0 if M

As follows,

	C1-T	C2	C3	C4-T	C5	C6
	Automobile	y	x10	x11	X11_new	X10*X11
14	Pacer	19.70	3375	A	1	3375
15	Babcat	20.30	2700	M	0	0
16	Granada	17.80	3890	A	1	3890
17	Eldorado	14.39	5290	A	1	5290
18	Emperial	14.89	5185	A	1	5185
19	Nova LN	17.80	3910	A	1	3910
20	Valiant	16.41	3660	A	1	3660
21	Starfire	23.54	3050	A	1	3050
22	Cordoba	21.47	4250	A	1	4250
23	Trans AM	16.59	3850	A	1	3850
24	Corolla E-5	31.90	2275	M	0	0
25	Astre	29.40	2150	M	0	0
26	Mark IV	13.27	5430	A	1	5430
27	Celica GT	23.90	2535	M	0	0
28	Charger SE	19.73	4370	A	1	4370
29	Cougar	13.90	4540	A	1	4540
30	Elite	13.27	4715	A	1	4715
31	Metador	13.77	4215	A	1	4215
32	Corvette	16.50	3660	A	1	3660

After changing values from A to 1 and M to 0, we have fitted the linear regression model as below,

Regression Equation

y = 37.92 - 0.004227 x10 - 3.72 X11_new

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
3.35421	73.64%	71.82%	66.71%

The coefficient table is given below,

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	37.92	2.56	14.80	0.000	
x10	-0.004227	0.000947	-4.47	0.000	2.25
X11_new	-3.72	1.98	-1.88	0.071	2.25

In the above table, we can see that the p-value for the regression model is 0.071 which is more than 0.05. Hence we can conclude that the type of transmission does not significantly affect the mileage performance.

b. Modify the model developed in part a to include an interaction between vehicle weight and the type of transmission. What conclusions can you draw about the effect of the type of transmission on gasoline mileage?

Solution:

Here we have to develop a model in part a which includes an interaction between the vehicle weight and the transmission type.

For this, We have to multiply X10 with X11_new for transmission and the result were stored in X10*X11

After this step,

The model is fitted using response variable y and continuous predictor as X10, X11_new and X10*X11
And we are getting the below regression model

Regression Equation

y = 58.11 - 0.01252 x10 - 28.66 X11_new + 0.00948 X10*X11

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	58.11	5.29	10.98	0.000	
x10	-0.01252	0.00214	-5.84	0.000	17.94
X11_new	-28.66	6.23	-4.60	0.000	34.73
X10*X11	0.00948	0.00229	4.14	0.000	82.84

Here we can see that the p-value is smaller than 0.05 which means that the type of transmission significantly affects the mileage performance.

Interpret the parameters in this model.
From the coefficient table,

1. If 1 unit of X10 changes then 0.01252 units decrease in y
2. If 1 unit of X11_new changes then 28.66 units decrease in y
3. If X10 increases then it will increase the significant effect of X10*X11 by 0.00948 units.

Problem-3: briefly describe your project data

The main objective behind this collaborative study is to gain knowledge and insights into the climate change data. For the project the predictor is temperature and the response variables are the various factors that affect the temperature like CO2, N2O,CH4 etc. The reason we have chosen temperature as the response variable is because the various factors in the atmosphere affect the temperature level and cause extreme events in the environment. The study of our project is to develop a linear regression model using the predictor variables given in the dataset and using those predictor variables we will predict the temperature.

Parameters in dataset:

The dataset contains data from May 1983 to December 2008. The available variables include:

Year: the observation year.

Month: the observation month.

Temp: the difference in degrees Celsius between the average global temperature in that period and a reference value. This data comes from the Climatic Research Unit at the University of East Anglia.

CO2, N2O, CH4, CFC.11, CFC.12: atmospheric concentrations of carbon dioxide (CO_2), nitrous oxide (N_2O), methane (CH_4), trichlorofluoromethane (CCl_3F; commonly referred to as CFC-11) and dichlorodifluoromethane (CCl_2F_2; commonly referred to as CFC-12), respectively. This data comes from the ESRL/NOAA Global Monitoring Division.

CO_2, N_2O and CH_4 are expressed in ppmv (parts per million by volume -- i.e., 397 ppmv of CO_2 means that CO_2 constitutes 397 millionths of the total volume of the atmosphere)

CFC.11 and CFC.12 are expressed in ppbv (parts per billion by volume).

Aerosols: the mean stratospheric aerosol optical depth at 550 nm. This variable is linked to volcanoes, as volcanic eruptions result in new particles being added to the atmosphere, which affect how much of the sun's energy is reflected back into space. This data is from the Godard Institute for Space Studies at NASA.

TSI: the total solar irradiance (TSI) in W/m^2 (the rate at which the sun's energy is deposited per unit area). Due to sunspots and other solar phenomena, the amount of energy that is given off by the sun varies substantially with time. This data is from the SOLARIS-HEPPA project website.

MEI: multivariate El Nino Southern Oscillation index (MEI), a measure of the strength of the El Nino/La Nina-Southern Oscillation (a weather effect in the Pacific Ocean that affects global temperatures). This data comes from the ESRL/NOAA Physical Sciences Division.