a) Denoting with $Y_i = mpg_i$ $\quad i = 1, \dots, 32$ (response variable)

and with x_{ij} = value of the j-th covariate on the i-th car

$j = 1, \dots, 9 \qquad i = 1, \dots, 32$

(with $x_{i1} = 1$ for all i, since the model includes the intercept)

The model can be written as

$$Y_i = \beta_1 + \beta_2 \underbrace{x_{i2}}_{wt_i} + \beta_3 \underbrace{x_{i3}}_{am_i} + \dots + \beta_9 \underbrace{x_{i9}}_{vs_i} + \varepsilon_i \qquad (\text{model A})$$

with $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, $\quad i = 1, \dots, 32$.

b) Sample space $\mathcal{Y} = \mathbb{R}^{32}$

Parameter space $\Theta = \mathbb{R}^9 \times (0, +\infty)$

c) 1. t-value of "hp"

this value corresponds to the observed test statistic for testing the hypothesis

$$\begin{cases} H_0: \beta_6 = 0 \\ H_1: \beta_6 \neq 0 \end{cases}$$

$$t^{obs} = \frac{\hat\beta_6 - 0}{\sqrt{\widehat{var}(\hat\beta_6)}} = \frac{\hat\beta_6}{\widehat{se}(\hat\beta_6)} = \frac{-0.0214}{0.0162} = -1.321$$

2. estimate of "qsec"

this value is $\hat\beta_{qsec} = \hat\beta_8$ (max. lik. estimate)

we can derive it inverting the formula used in point 1.

$$t^{obs} = \frac{\hat\beta_8}{\widehat{se}(\hat\beta_8)} \Rightarrow \hat\beta_8 = t^{obs} \cdot \widehat{se}(\hat\beta_8)$$
$$= 1.229 \cdot 0.6587 = 0.8095$$

3. Pr(>|t|) of "vs"

this is the pvalue of the test of significance of $\beta_{vs} = \beta_9$

$$\begin{cases} H_0: \beta_9 = 0 \\ H_1: \beta_9 \neq 0 \end{cases}$$

the test statistic (pivotal quantity) is

$$T = \frac{\hat\beta_9}{\sqrt{\widehat{var}(\hat\beta_9)}} \overset{H_0}{\sim} t_{n-p} = t_{32-9} = t_{23}$$

the pvalue is
$$\alpha^{obs} = \mathbb{P}_{H_0}(|T| > |t^{obs}|)$$
$$= 2\, \mathbb{P}_{H_0}(T > |t^{obs}|)$$
$$= 2\, \mathbb{P}_{H_0}(T > 0.183)$$

where $T \sim t_{23}$

Student's t with 23 dof.

d) We want to test

$$\begin{cases} H_0: \beta_2 = \beta_3 = \dots = \beta_8 = \beta_9 = 0 \\ H_1: \text{at least one of } \beta_j \text{ is } \neq 0 \quad (j = 2, \dots, 9) \end{cases}$$

we use the statistic

$$F = \frac{\tilde\sigma^2 - \hat\sigma^2}{\hat\sigma^2} \cdot \frac{n-p}{p-1} \overset{H_0}{\sim} F_{p-1, n-p}$$

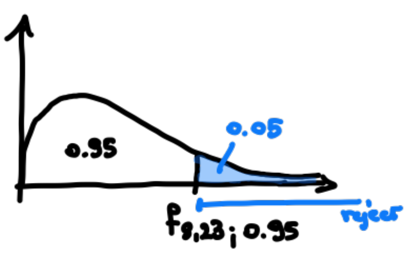
with $\tilde\sigma^2$ estimate of σ^2 under the null model $\quad \tilde\sigma^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar y)^2 = \frac{1}{n} SST \quad (H_0)$

$\hat\sigma^2$ estimate under the full model "A" $\quad \hat\sigma^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat y_i)^2 = \frac{1}{n} SSE \quad (H_1)$

$$F = \frac{SST - SSE}{SSE} \cdot \frac{23}{8} = \frac{SSR/SST}{SSE/SST} \cdot \frac{23}{8} = \frac{R^2}{1-R^2} \cdot \frac{23}{8}$$

with the data I get $f^{obs} = \frac{0.8678}{1 - 0.8678} \cdot \frac{23}{8} = 18.87$

at a significance level of 5%: I reject H_0 if $f^{obs} > f_{8,23;0.95}$ $\;(= 2.3748)$



hence I reject H_0.

e) model B

$$Y_i = \beta_1 + \beta_2 \underbrace{x_{i2}}_{wt_i} + \beta_3 \underbrace{x_{i3}}_{am_i} + \varepsilon_i$$

We can use a test for comparing nested models, since model B can be obtained as a restriction of model A. Specifically, we need to test

$$\begin{cases} H_0: \beta_4 = \beta_5 = \dots = \beta_8 = \beta_9 = 0 \\ H_1: \text{at least one of } \beta_j \text{ is } \neq 0 \quad (j = 4, \dots, 9) \end{cases}$$

the test statistic is

$$F = \frac{\tilde\sigma^2 - \hat\sigma^2}{\hat\sigma^2} \cdot \frac{n-p}{p-p_0} \overset{H_0}{\sim} F_{p-p_0, n-p}$$

where

$\tilde\sigma^2$ estimate of σ^2 under model B (H_0) $\quad \tilde\sigma^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat y_i^B)^2 = \frac{SSE_B}{n}$

$\hat\sigma^2$ estimate under the full model A $\quad \hat\sigma^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat y_i^A)^2 = \frac{SSE_A}{n}$

$n = 32$

$p = 9 \quad p_0 = 3$

$$SSE_A = (n-p) \cdot (\text{residual SE})^2 = 23 \cdot 2.544^2 = 148.85$$
$$SSE_B = (n-p_0) \cdot (\text{residual SE})^2 = 29 \cdot 3.098^2 = 278.33$$

$$f^{obs} = \frac{278.33 - 148.85}{148.85} \cdot \frac{23}{6} = 3.334$$

using a 5% significance level, I reject H_0 if $f^{obs} > f_{6,23;0.95}$ $\;(= 2.527)$

Hence, I reject H_0: I prefer model A.

using a 1% significance level, I reject H_0 if $f^{obs} > f_{6,23;0.99}$ $\;(= 3.71)$

Here I do not reject H_0

f) introducing the interaction we obtain the model

$$Y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \varepsilon_i$$

where $\quad x_{i2} = wt_i$ (weight of car i)

$$x_{i3} = am_i = \begin{cases} 1 & \text{if car } i \text{ has manual transmission} \\ 0 & \text{if car } i \text{ has automatic transmission} \end{cases}$$

$$x_{i4} = x_{i2} \cdot x_{i3} = \begin{cases} x_{i2} = wt_i & \text{if } am_i = 1 \\ 0 & \text{if } am_i = 0 \end{cases} \quad (\text{weight of car } i \text{ if manual})$$

interpretation of parameters

- consider a car with automatic transmission ($x_{i3} = 0$, $x_{i4} = 0$)

  the mean consumption is

  $\mu_i = \beta_1 + \beta_2 x_{i2}$

- for a car with manual transmission ($x_{i3} = 1$, $x_{i4} = x_{i2}$)

  $\mu_i = \beta_1 + \beta_2 x_{i2} + \beta_3 + \beta_4 x_{i2}$

  $= (\beta_1 + \beta_3) + (\beta_2 + \beta_4) x_{i2}$

β_1 is the intercept for cars with an automatic transmission

$\beta_1 + \beta_3$ is the intercept for cars with a manual transmission

β_2 is the effect on the mean consumption of increasing the weight of the car of 1 unit, for cars with an automatic transmission

$\beta_2 + \beta_4$ is the effect on the mean consumption of increasing the weight of the car of 1 unit, for cars with a manual transmission