12265092

Question 1

a) Calculations done in Excel sheet attached

Source	SS	df	MS	**Number of obs**	=	**2298**
				F(3, 2294)	=	**554**
Model	190.873336	3.00	63.624445400	**Prob > F**	=	**0.000**
Residual	263.547174	2294.00	0.114885429	**R-squared**	=	**0.420**
				Adj R-squared	=	**0.419**
Total	454.420510	2297.00	0.197832177	**Root MSE**	=	**0.339**
lfare	Coef.	Std. Err.	t	P>|t|	[95% conf. interval]	
ldist	0.4544197	0.128003	**3.550**	**0.0004180**	**0.204**	**0.705**
lpassen	-0.0678697	0.0084094	**-8.071**	**0.000**	**-0.084**	**-0.051**
bmktshr	0.3092456	0.0425167	**7.274**	**0.000**	**0.226**	**0.393**
_cons	2.241328	0.1218621	**18.392**	**0.000**	**2.002**	**2.480**

formulas

$$F(3,2294) = \frac{\text{MS of Model}}{\text{MS of Residual}}.$$

$$R^2 = 1 - \frac{SSR}{SST}$$

$$\text{Adj } R^2 = 1 - \frac{n-1}{n-k-1} \cdot \frac{SSR}{SST}.$$

$$\text{Root MSE} = \text{sqrt}(\text{MS of Residuals})$$

p-values = Using 2tail tests (Excel)

b) Interpret the Coefficients on ldist and bmktshr in English sentence

Coefficient of ldist(0.4544197) is the elasticity of lfare with respect to ldist
Coefficient of bmktshr(0.3092456) is the elasticity of lfare with respect to bmktshr

ldist: When ldist increases by 1%, lfare increases by 0.4544197%, holding all other independent variables fixed.

bmktshr: When bmktshr increases by 1%, lfare increases by 0.3092456%, holding all other independent variables fixed.

c) Test null hypothesis that the coefficient of bmktshr is equal to zero

1(c) Test $H_0: \beta_{bmktshr} = 0$

2-tailed t-test

$$t \text{ statistic} = \frac{0.3092456}{0.0425167} = 7.2735$$

p-value = 0.00 (calculated in 1a)

P-value (0.00) is less than the p-value for 5% confidence level. $0 < 0.05$

Hence, $H_0: \beta_{bmktshr} = 0$ can be rejected.
However, the alternate hypothesis that coefficient of "bmktshr" is not zero cannot be rejected.

d) Interpret the R-squared and the overall F-statistic

R-squared

1(d) $R^2 = 0.420$

R^2 is the percentage of the dependent variable variation that a model can explain.

$$R^2 = \frac{\text{variance exp. by model}}{\text{Total var.}}$$

R^2 of 42% indicates that 42% variance of l airfare, of the dependent variable can be explaind by the model with independent variables ldisT, bmkTshr, lpassen. where ldist is log distance bmkshr is market share and l passn is no. of passengers.

F-statistic

f - statistic

$$f_{stat} = \frac{R^2/k}{(1-R^2)/(n-k-1)}$$

$$= \frac{0.420/3}{(1-0.42)/(2298-3-1)}$$

$$= \frac{0.42}{0.58} \times \frac{2294}{3}$$

$$\boxed{f_{value} \simeq 554} = \frac{\text{MS of } \cancel{\text{Residuals}} \text{ Model}}{\text{MS of Residuals}}$$

P-value associated is 0.00, this when compared with 5% significance level (p 0.05). We can reject the null hypothesis that the independent variables are insignificant in determining airfare.

We can however not reject the alternate hypothesis that the variables l distance, bmktshr, ~~l~~ l no. of passengers are significant in determining l airfare.

Question 2

a) First, test the hypothesis that $H0$: $\beta\beta 0 = 0$ against the two-sided alternative. Then test $H0$: $\beta 1 = 1$ against the two-sided alternative. What do you conclude?

Q2

$$\text{price} = \beta_0 + \beta_1 \text{assess} + u$$

$$\widehat{\text{price}} = -14.47 + 0.976\ \text{assess} + u$$
$$\qquad\quad (16.27) \qquad\quad (0.049)$$

$n = 88 \qquad SSR = 165644.51 \qquad R^2 = 0.820$

ⓐ $\Rightarrow$ $H_0: \beta_0 = 0$ against 2 sided alternative

$\hat{\beta}_0 = -14.47 \qquad \beta_0 = 0 \qquad s.e(\hat{\beta}_0) = 16.27$

$$|tstat| = \left|\frac{-14.47 - 0}{16.27}\right| = 0.8894$$

The critical value at 5% significance level, two sided, d.o.f = 86 $(88-1-1)$ is 1.98. As the absolute tstat $0.8894 < 1.98$ the null hypothesis H_0 cannot be rejected.

$\Rightarrow$ $H_0: \beta_1 = 1$ against 2 sided alternative.

$\hat{\beta}_1 = 0.976 \qquad \beta_1 = 1 \qquad s.e(\hat{\beta}_1) = 0.049.$

$$|tstat| = \left|\frac{0.976 - 1}{0.049}\right| = |-0.489| = 0.489.$$

Critical value at 5% level, two sided, d.o.f = 86 is 1.98. As the absolute value of tstatistic is very less compared to 1.98, we fail to reject the null hypothesis $H_0: \beta_1 = 1$

$\Rightarrow$ As we cannot reject both the H_0's, the variable assess (assessment) is rational and significant in determining price.

b) Joint hypothesis, B0 = 0 and B1 = 1

2 (b) Joint hypothesis $\beta_0 = 0$ $\beta_1 = 1$, perform f test.

Given $SSR_{ur} = 165644.51$ $n = 88$

$SSR_r = 209448.99$

$$f = \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)}$$

$q = df_r - df_{ur} = 1$

$n-k-1 = df_{ur} = 88-1-1 = 86.$

$$f = \frac{(209448.99 - 165644.51)/1}{(165644.51)/86}.$$

$f_{stat} = 22.742$

Critical value from f distribution for 5% level $q=1$ and $n-k-1 = 86$ is $\sim 3.96.$ (3.9519)

As the value of fstat calculated (22.742) is very high than the critical value, we can reject the null hypothesis Ho. As Ho is rejected we can say that ~~joint hypothesis~~ both the β_0 and β_1 (i.e) intercept and the coeff., assess (assessment) is ~~jointly~~ significant statistically at the 5% level.

Also it is effective to use t statistic over f statistic as t is flexible for single variable hypothesis and can be used against 1 sided alternatives

c) Test H0: B2 = 0 , B3 = 0, B4 = 0

2 ⓒ

$price = \beta_0 + \beta_1 assess + \beta_2 lotsize + \beta_3 sqrft + \beta_4 bdrms + u.$

Test $H_0: \beta_2 = 0 \quad \beta_3 = 0 \quad \beta_4 = 0$

Given $R^2_{ur} = 0.829$

$R^2_r = 0.82$ (from 2a)

$$f = \frac{(R^2_{ur} - R^2_r)/q}{(1 - R^2_{ur})/(n-k-1)}.$$

$q = df_r - df_{ur} = 3 \quad (88-2-88+5)$

$n-k-1 = df_{ur} = 83 \quad (88-4-1)$

$$f = \frac{(0.829 - 0.82)/3}{(1 - 0.829)/83} = 1.4561$$

Critical value from f distribution for 5% significance level, $q = 3$, $n-k-1 = 83$ is ≈ 2.7146

As the f value calculated is less than critical value, we cannot reject Ho the null hypothesis. Thus, the variables lotsize, sqrft, bdrms are jointly insignificant at the 5% level.

d) If the variance of price changes with assess, lotsize, sqrft, or bdrms, what can you say about the F-test from part (c)

2 (d) In 2c we saw that the variables lotsize, sqrft, bdrms are jointly insignificant at the 5% significance level.

Now if variance of price changes with all variables, we can say this is attributed to SSR. This is because as per MLR 5, it says that variance of y given x, does not depend on the values of the independent variables.

But this does not make sense, as according to homoskedasticity, it is required that variance of u does not depend on independent variables.

So when we are saying variance changes with x variables, then our assumption of homoskedasticity does not apply. That means the f-test from the part 2c does not validate the hypothesis. We need to then test each variable seperately to see the significance of the variable in the model.

Question 3

a) Question 3a

Question 3

(a) The Y_i's here are $(x_i-\bar{x})^2$ and mean of each is equal to $Var(x_i)$, so by LLN the denominator converges to $Var(x_i)$

(b) $E[(x_i-\bar{x})u_i] = E[x_i u_i] - \bar{x}E[u_i]$

$= E[x_i E[u_i|x]] - \bar{x}E[E[u_i|x]]$

$= E[x_i \cdot E[u_i]] - \bar{x}E[E[u_i]]$

$= E[x_i \cdot 0] - \bar{x}E[0]$

$= 0$

from above, Y_i i.e $(x_i-\bar{x})u_i$ have mean zero.

b) 3b

(b) $E[(x_i-\bar{x})u_i] = E[x_i u_i] - \bar{x}E[u_i]$

$= E[x_i E[u_i|x]] - \bar{x}E[E[u_i|x]]$

$= E[x_i \cdot E[u_i]] - \bar{x}E[E[u_i]]$

$= E[x_i \cdot 0] - \bar{x}E[0]$

$= 0$

from above, Y_i i.e $(x_i-\bar{x})u_i$ have mean zero.

c) 3c

(c) $Var[(x_i-\bar{x})u_i] = Var(E[(x_i-\bar{x})u_i|x]) + E[Var((x_i-\bar{x})u_i|x)]$

$= Var[E[x_iu_i - \bar{x}u_i]] + E[Var((x_i-\bar{x})u_i|x)]$

$= Var[E[x_iu_i] - \bar{x}E(u_i)] + Var((x_i-\bar{x})u_1 + (x_2-\bar{x})u_2 + \ldots (x_n-\bar{x})u_n) \Rightarrow E[(x_i-\bar{x})^2]\cdot Var(u)$

from 3b

$= 0 + (x_i-\bar{x})^2 Var(u_1) + (x_2-\bar{x})Var(u_2) \ldots (x_n-\bar{x})^2 \cdot Var(u_n)$

$= 0 + E[(x_i-\bar{x})^2] * Var(u_i)$

$= \sigma^2 \cdot \sum_{i=1}^{n}(x_i-\bar{x})^2$

$= \sigma^2 Var(x_i)$.

We used homoskedasticity $Var(u|x) = Var(u) = \sigma^2$

d) 3d

(d) we know

$$\sqrt{n}\left(\hat{\beta}_1 - \beta_1\right) = \sqrt{n} \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})u_i}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

from 2b. we found $E(\text{numerator}) = 0$

from 2c we found variance of numerators v_i terms is equal to $\sigma^2 \operatorname{Var}(x_i)$.

$\Rightarrow$ Now using the Central limit theorem, we can say that, the distribution formed σZ converges to $N(0, \sigma^2 \operatorname{Var}(x))$

$\Rightarrow$ Denominator term $a_n = \left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{-1}$

converges to $\simeq \operatorname{Var}(x_i)$

Now we have $\frac{\text{numerator}}{\text{denominator}} = \frac{N(0, \sigma^2 \operatorname{Var}(x))}{\operatorname{Var}(x_i)}$

$$= N\left(0, \frac{\sigma^2}{\operatorname{Var}(x_i)}\right)$$

Continued…

Given fact that, If a random variable X is such that $\sqrt{n}X$ converges in distribution to $N(0, v)$, then for an a_n such that $\text{plim } a_n = a$, $\sqrt{n}\, a_n X$ converges in distribution to $N(0, a^2 v)$.

In our case $\sqrt{n}X$ = numerator converges to $N(0, \sigma^2 Var(x_i))$

then for a_n (denominator) where $\text{plim } a_n = a$.

The $\sqrt{n}\, a_n X$ (i.e) $\sqrt{n}(\hat{\beta}_1 - \beta_1)$ converges in distribution to $N\left(0, \frac{\sigma^2}{Var(x_i)}\right)$

Thus we can say

$\sqrt{n}(\hat{\beta}_1 - \beta_1)$ is asymptotically distributed $N\left(0, \frac{\sigma^2}{Var(x_i)}\right)$