```
                                                                  ___  ____  ____  ____  ____(R)
                                                                 /__    /   ____/   /   ____/
                                                                ___/   /   /___/   /   /___/
                                                                  Statistics/Data analysis

                                                                     User: Airline Route analysis
                                                                  Project: Route Analysis
```

```
      name:  <unnamed>
       log:  [redacted]
  log type:  smcl
 opened on:  9 Dec 2025, 16:19:26

 1 . egen zHHI = std(HHI)

 2 . gen dist2 = Distance^2

 3 . gen pass2 = Passengers^2

 4 . gen lpassen = log(Passengers)

 5 . gen lfuelcost = log(FuelCost)
   (6,437 missing values generated)

 6 . gen lnfare = log(Fare)

 7 . gen lndist = log(Distance)

 8 . gen lcHHI = HHI * Lowcost

 9 . gen lczHHI = zHHI * Lowcost

10 . sum Distance Caputil Lowcost Passengers Fare HHI zHHI lndist FuelCost lnfare lpassen
```

Variable	Obs	Mean	Std. dev.	Min	Max
Distance	222,252	919.2392	651.6014	27	4983
Caputil	222,220	.6639203	.1450958	0	1.007576
Lowcost	222,252	.2598312	.4385428	0	1
Passengers	222,252	1207.767	1518.202	2	20372
Fare	222,252	166.4496	75.01857	22.1105	1473.102
HHI	222,252	.752702	.2455447	.1446151	1
zHHI	222,252	-8.32e-09	1	-2.476481	1.00714
lndist	222,252	6.573652	.7359745	3.295837	8.513787
FuelCost	215,815	225.524	160.9346	.59158	970.298
lnfare	222,252	5.007492	.4790818	3.096053	7.295126
lpassen	222,252	6.481173	1.254088	.6931472	9.921917

```
11 . corr Distance Caputil Lowcost Passengers Fare HHI zHHI lndist FuelCost lnfare lpassen
   (obs=215,784)
```

	Distance	Caputil	Lowcost	Passen~s	Fare	HHI	zHHI	lndist	FuelCost	lnfare
Distance	1.0000									
Caputil	0.3507	1.0000								
Lowcost	0.0770	0.1504	1.0000							
Passengers	-0.1029	0.1739	0.2302	1.0000						
Fare	0.4918	0.0403	-0.1869	-0.2520	1.0000					
HHI	-0.1133	-0.0546	-0.3402	-0.1868	0.1859	1.0000				
zHHI	-0.1133	-0.0546	-0.3402	-0.1868	0.1859	1.0000	1.0000			
lndist	0.9144	0.3875	0.1011	-0.1186	0.4743	-0.1068	-0.1068	1.0000		
FuelCost	0.1211	0.1982	-0.0325	0.0213	0.2038	0.0033	0.0033	0.1321	1.0000	
lnfare	0.4916	0.0817	-0.1614	-0.2915	0.9586	0.1881	0.1881	0.5204	0.2305	1.0000
lpassen	-0.0079	0.2765	0.2183	0.7264	-0.2035	-0.2142	-0.2142	0.0221	0.0225	-0.2132

	lpassen
lpassen	1.0000

```
12 . scatter Fare Distance

13 .
14 . scatter Fare Caputil

15 .
16 . ...

17 . scatter Fare Distance
```

[redacted]

```
19 . scatter Fare Caputil

20 .
```

```
      name:  <unnamed>
```

[redacted]

```
  log type:  smcl
 opened on:  9 Dec 2025, 17:01:15
```

[redacted]

```
23 . scatter Fare Distance

24 . scatter Fare Lowwcost
variable Lowwcost not found
r(111);

25 . scatter Fare Lowcost

26 . scatter Fare Passengers

27 . scatter Fare HHI

28 . scatter Fare lndist
variable lndist not found
r(111);

29 . gen lndist = log(Distance)

30 . gen lcHHI = HHI * Lowcost

31 . gen lczHHI = zHHI * Lowcost
zHHI not found
r(111);

32 . egen zHHI = std(HHI)

33 . * 1. Standardized HHI

34 .
```

```
35 . egen zHHI = std(HHI)
   variable zHHI already defined
   r(110);

36 .
37 .
38 .
39 . * 2. Distance squared

40 .
41 . gen dist2 = Distance^2

42 .
43 .
44 .
45 . * 3. Passengers squared

46 .
47 . gen pass2 = Passengers^2

48 .
49 .
50 .
51 . * 4. Log of passengers

52 .
53 . gen lpassen = log(Passengers)

54 .
55 .
56 .
57 . * 5. Log of fuel cost

58 .
59 . gen lfuelcost = log(FuelCost)
   (6,437 missing values generated)

60 .
61 .
62 .
63 . * 6. Log of fare

64 .
65 . gen lnfare = log(Fare)

66 .
67 .
68 .
69 . * 7. Log of distance

70 .
71 . gen lndist = log(Distance)
   variable lndist already defined
   r(110);
```

```
72 .
73 .
74 .
75 . * 8. Interaction: HHI × Lowcost

76 .
77 . gen lcHHI = HHI * Lowcost
   variable lcHHI already defined
   r(110);

78 .
79 .
80 .
81 . * 9. Interaction: zHHI × Lowcost

82 .
83 . gen lczHHI = zHHI * Lowcost

84 . scatter Fare lndist

85 . scatter Fare FuelCost

86 . scatter Fare lpassen

87 . scatter Fare zHHI

88 . hist Distance, percent
   (bin=53, start=27, width=93.509434)

89 . hist Passengers, percent
   (bin=53, start=2, width=384.33962)

90 . hist Fare, percent
   (bin=53, start=22.110498, width=27.377204)

91 . hist HHI, percent
   (bin=53, start=.14461508, width=.01613934)

92 . hist lndist, percent
   (bin=53, start=3.2958369, width=.09845189)

93 . hist lnfare, percent
   (bin=53, start=3.0960526, width=.0792278)

94 . hist FuelCost, percent
   (bin=53, start=.59157997, width=18.296347)

95 . hist lpassen, percent
   (bin=53, start=.69314718, width=.17412773)

96 . hist zHHI, percent
   (bin=53, start=-2.4764812, width=.06572871)

97 . scatter lnfare Distance
```

```
 98 . scatter lnfare Caputil

 99 . scatter lnfare Passengers

100 . scatter lnfare HHI

101 . scatter lnfare lndist

102 . scatter lnfare FuelCost

103 . scatter lnfare lpassen

104 . scatter lnfare zHHI

105 . log close
      name:  <unnamed>
       log:  [redacted]
  log type:  smcl
 closed on:   9 Dec 2025, 18:40:40
-------------------------------------------------------------------------------

-------------------------------------------------------------------------------
      name:  <unnamed>
       log:  [redacted]
  log type:  smcl
 opened on:  10 Dec 2025, 18:06:24

106 . desc

[redacted]

  Observations:       222,252
     Variables:            17                  9 Dec 2025 18:45
-------------------------------------------------------------------------------
Variable      Storage   Display    Value
    name         type    format    label      Variable label
-------------------------------------------------------------------------------
year            float   %8.0g                 (mean) year
Distance        float   %8.0g                 Route distance (miles)
Caputil         float   %9.0g                 Capacity utilization
Lowcost         float   %9.0g                 Low-Cost Carrier
Passengers      float   %9.0g                 Total passengers on route
Fare            float   %9.0g                 Mean one-way fare ($)
HHI             float   %9.0g                 Market concentration (HHI)
FuelCost        float   %9.0g                 Fuel cost ($1,000s)
lndist          float   %9.0g
lcHHI           float   %9.0g
zHHI            float   %9.0g                 Standardized values of HHI
dist2           float   %9.0g
pass2           float   %9.0g
lpassen         float   %9.0g
lfuelcost       float   %9.0g
lnfare          float   %9.0g
lczHHI          float   %9.0g
-------------------------------------------------------------------------------
Sorted by:
```

```
107 . browse

108 . sum
```

Variable	Obs	Mean	Std. dev.	Min	Max
year	222,252	1999.296	3.948386	1993	2006
Distance	222,252	919.2392	651.6014	27	4983
Caputil	222,220	.6639203	.1450958	0	1.007576
Lowcost	222,252	.2598312	.4385428	0	1
Passengers	222,252	1207.767	1518.202	2	20372
Fare	222,252	166.4496	75.01857	22.1105	1473.102
HHI	222,252	.752702	.2455447	.1446151	1
FuelCost	215,815	225.524	160.9346	.59158	970.298
lndist	222,252	6.573652	.7359745	3.295837	8.513787
lcHHI	222,252	.159394	.2951519	0	1
zHHI	222,252	-8.32e-09	1	-2.476481	1.00714
dist2	222,252	1269583	1907579	729	2.48e+07
pass2	222,252	3763629	1.33e+07	4	4.15e+08
lpassen	222,252	6.481173	1.254088	.6931472	9.921917
lfuelcost	215,815	5.083544	.9810632	-.5249584	6.877603
lnfare	222,252	5.007492	.4790818	3.096053	7.295126
lczHHI	222,252	-.1473519	.5534725	-2.476481	1.00714

```
109 . describe fare distance passengers fuelcost caputil lowcost hhi lnfare zHHI dist2 pass2 lczHHI
variable fare not found
r(111);

110 . describe Fare distance passengers fuelcost caputil lowcost hhi lnfare zHHI dist2 pass2 lczHHI
variable distance not found
r(111);

111 . describe Fare Distance Passengers FuelCost Caputil Lowcost HHI lndist lHHI zHHI lnfare dist2 pass2 lczHHI
variable lHHI not found
r(111);

112 . reg lnfare Caputil Lowcost FuelCost zHHI Distance dist2 Passengers pass2 lczHHI
```

Source	SS	df	MS		
Model	19836.7297	9	2204.08108	Number of obs	215,784
Residual	28811.759	215,774	.133527483	F(9, 215774)	16506.57
Total	48648.4887	215,783	.22545098	Prob > F	0.0000
				R-squared	0.4078
				Adj R-squared	0.4077
				Root MSE	.36541

lnfare	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
Caputil	-.3294948	.0062468	-52.75	0.000	-.3417384	-.3172511
Lowcost	-.1231046	.0021352	-57.65	0.000	-.1272895	-.1189196
FuelCost	.0005325	5.02e-06	106.04	0.000	.0005226	.0005423
zHHI	.1054736	.0009949	106.02	0.000	.1035237	.1074236
Distance	.0006445	3.58e-06	179.85	0.000	.0006375	.0006516
dist2	-9.89e-08	1.18e-09	-83.98	0.000	-1.01e-07	-9.66e-08
Passengers	-.0000719	1.14e-06	-62.97	0.000	-.0000741	-.0000696
pass2	2.76e-09	1.24e-10	22.17	0.000	2.51e-09	3.00e-09
lczHHI	-.0765273	.0019151	-39.96	0.000	-.0802808	-.0727737
_cons	4.745726	.0038715	1225.82	0.000	4.738138	4.753314

```
113 . test Distance dist2

 ( 1)  Distance = 0
 ( 2)  dist2 = 0

       F(  2,215774) =41789.72
            Prob > F =    0.0000

114 . test Passengers pass2

 ( 1)  Passengers = 0
 ( 2)  pass2 = 0
       Constraint 2 dropped

       F(  1,215774) = 3965.00
            Prob > F =    0.0000

115 . reg lnfare Caputil Lowcost FuelCost zHHI Distance dist2 Passengers pass2 lczHHI
```

Source	SS	df	MS	Number of obs	=	215,784
				F(9, 215774)	=	16506.57
Model	19836.7297	9	2204.08108	Prob > F	=	0.0000
Residual	28811.759	215,774	.133527483	R-squared	=	0.4078
				Adj R-squared	=	0.4077
Total	48648.4887	215,783	.22545098	Root MSE	=	.36541

lnfare	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
Caputil	-.3294948	.0062468	-52.75	0.000	-.3417384	-.3172511
Lowcost	-.1231046	.0021352	-57.65	0.000	-.1272895	-.1189196
FuelCost	.0005325	5.02e-06	106.04	0.000	.0005226	.0005423
zHHI	.1054736	.0009949	106.02	0.000	.1035237	.1074236
Distance	.0006445	3.58e-06	179.85	0.000	.0006375	.0006516
dist2	-9.89e-08	1.18e-09	-83.98	0.000	-1.01e-07	-9.66e-08
Passengers	-.0000719	1.14e-06	-62.97	0.000	-.0000741	-.0000696
pass2	2.76e-09	1.24e-10	22.17	0.000	2.51e-09	3.00e-09
lczHHI	-.0765273	.0019151	-39.96	0.000	-.0802808	-.0727737
_cons	4.745726	.0038715	1225.82	0.000	4.738138	4.753314

```
116 . estat imtest, white

    White's test
    H0: Homoskedasticity
    Ha: Unrestricted heteroskedasticity

       chi2(48) = 36267.79
    Prob > chi2 =   0.0000

    Cameron & Trivedi's decomposition of IM-test
```

Source	chi2	df	p
Heteroskedasticity	36267.79	48	0.0000
Skewness	3567.99	9	0.0000
Kurtosis	672.53	1	0.0000
Total	40508.30	58	0.0000

```
117 . log close
      name:  <unnamed>
       log:  [redacted]
  log type:  smcl
 closed on:  10 Dec 2025, 19:05:10
-------------------------------------------------------------------------------

-------------------------------------------------------------------------------
      name:  <unnamed>
       log:  [redacted]
  log type:  smcl
 opened on:  11 Dec 2025, 19:19:24

[redacted]

120 .
[redacted]
r(604);

end of do-file

r(604);

122 . summarize Distance Passengers Fare HHI, detail

                    Route distance (miles)
-------------------------------------------------------------
      Percentiles      Smallest
 1%           110              27
 5%           209              27
10%           260              27       Obs             222,252
25%           418              27       Sum of wgt.     222,252

50%           762                       Mean           919.2392
                       Largest          Std. dev.      651.6014
75%          1185            4962
90%          1870            4983       Variance       424584.4
95%          2342            4983       Skewness       1.380149
99%          2677            4983       Kurtosis       5.436565

                   Total passengers on route
-------------------------------------------------------------
      Percentiles      Smallest
 1%             5               2
 5%           110               2
10%           161               2       Obs             222,252
25%           327               2       Sum of wgt.     222,252

50%           704                       Mean           1207.767
                       Largest          Std. dev.      1518.202
75%          1484           18180
90%          2755           18278       Variance        2304938
95%          3998           18298       Skewness       3.451807
99%          7605           20372       Kurtosis       20.83419

                    Mean one-way fare ($)
-------------------------------------------------------------
      Percentiles      Smallest
 1%     45.57948         22.1105
 5%     62.96188        22.29661
10%     75.88485        22.29661       Obs             222,252
25%     109.6078        22.35519       Sum of wgt.     222,252
```

50%	**156.1667**		Mean	**166.4496**
		Largest	Std. dev.	**75.01857**
75%	**215.6842**	**838.875**		
90%	**266.1687**	**1286.269**	Variance	**5627.786**
95%	**298.6802**	**1452.059**	Skewness	**.794162**
99%	**374.8478**	**1473.102**	Kurtosis	**4.770119**

Market concentration (HHI)

	Percentiles	Smallest		
1%	**.2641728**	**.1446151**		
5%	**.3561649**	**.1446151**		
10%	**.4233422**	**.1446151**	Obs	**222,252**
25%	**.5179117**	**.1446151**	Sum of wgt.	**222,252**
50%	**.801401**		Mean	**.752702**
		Largest	Std. dev.	**.2455447**
75%	**1**	**1**		
90%	**1**	**1**	Variance	**.0602922**
95%	**1**	**1**	Skewness	**-.3111749**
99%	**1**	**1**	Kurtosis	**1.537362**

```
123 . misstable summarize
```

			Obs<.			
Variable	Obs=.	Obs>.	Obs<.	Unique values	Min	Max
Caputil	**32**		**222,220**	>500	0	1.007576
FuelCost	**6,437**		**215,815**	>500	.59158	970.298
lfuelcost	**6,437**		**215,815**	>500	-.5249584	6.877603

```
124 . kdensity Fare, normal

125 .
126 . kdensity Passengers, normal

127 .
128 . kdensity Distance, normal

129 . vif
    not appropriate after regress, nocons;
    use option uncentered to get uncentered VIFs
    r(301);

130 . pwcorr Distance Passengers HHI Fare FuelCost, sig star(0.05)
```

	Distance	Passen~s	HHI	Fare	FuelCost
Distance	**1.0000**				
Passengers	**-0.0984*** **0.0000**	**1.0000**			
HHI	**-0.1070*** **0.0000**	**-0.1855*** **0.0000**	**1.0000**		
Fare	**0.4929*** **0.0000**	**-0.2411*** **0.0000**	**0.1855*** **0.0000**	**1.0000**	
FuelCost	**0.1209*** **0.0000**	**0.0213*** **0.0000**	**0.0033** **0.1269**	**0.2037*** **0.0000**	**1.0000**

```
131 . estat ovtest
    last estimates not found
    r(301);

132 . predict uhat, resid
    last estimates not found
    r(301);

133 .
134 . scatter uhat Distance
    variable uhat not found
    r(111);

135 .
136 . scatter uhat Passengers
    variable uhat not found
    r(111);

137 .
138 . rvfplot
    last estimates not found
    r(301);

139 . reg lnfare Caputil Lowcost FuelCost zHHI Distance dist2 Passengers pass2 lcHHI
```

Source	SS	df	MS		
Model	19836.7297	9	2204.08108	Number of obs	= 215,784
Residual	28811.759	215,774	.133527483	F(9, 215774)	= 16506.57
				Prob > F	= 0.0000
Total	48648.4887	215,783	.22545098	R-squared	= 0.4078
				Adj R-squared	= 0.4077
				Root MSE	= .36541

lnfare	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
Caputil	-.3294948	.0062468	-52.75	0.000	-.3417384	-.3172511
Lowcost	.111485	.0055022	20.26	0.000	.1007007	.1222692
FuelCost	.0005325	5.02e-06	106.04	0.000	.0005226	.0005423
zHHI	.1054736	.0009949	106.02	0.000	.1035237	.1074236
Distance	.0006445	3.58e-06	179.85	0.000	.0006375	.0006516
dist2	-9.89e-08	1.18e-09	-83.98	0.000	-1.01e-07	-9.66e-08
Passengers	-.0000719	1.14e-06	-62.97	0.000	-.0000741	-.0000696
pass2	2.76e-09	1.24e-10	22.17	0.000	2.51e-09	3.00e-09
lcHHI	-.3116632	.0077994	-39.96	0.000	-.3269498	-.2963766
_cons	4.745726	.0038715	1225.82	0.000	4.738138	4.753314

```
140 . predict uhat, resid
    (6,468 missing values generated)

141 . rvfplot

142 .
143 . scatter uhat Distance
```

```
144 .
145 . qnorm uhat

146 .
147 . swilk uhat
```

Shapiro–Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
uhat	**215,784**	**0.99776**	**112.834**	**13.352**	**0.00000**

Note: The normal approximation to the sampling distribution of W'
is valid for 4<=n<=2000.

```
148 . log close
      name:  <unnamed>
       log:  [redacted]
  log type:  smcl
 closed on:  11 Dec 2025, 19:28:25
```