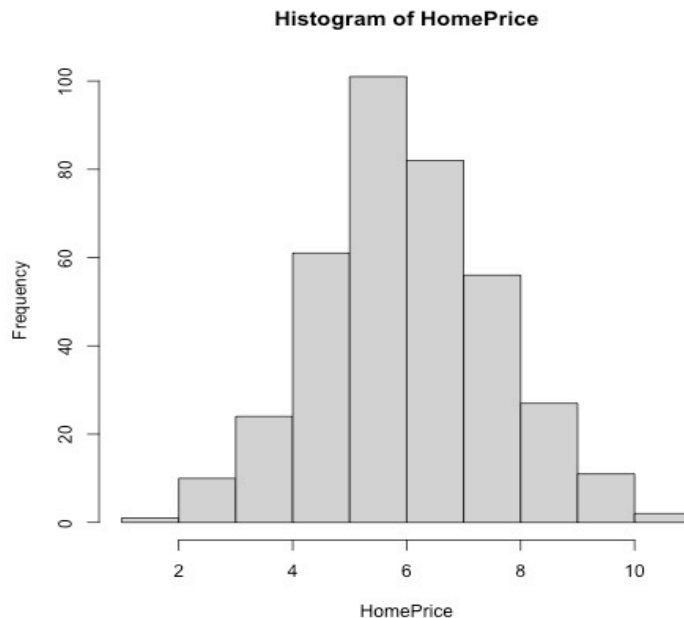


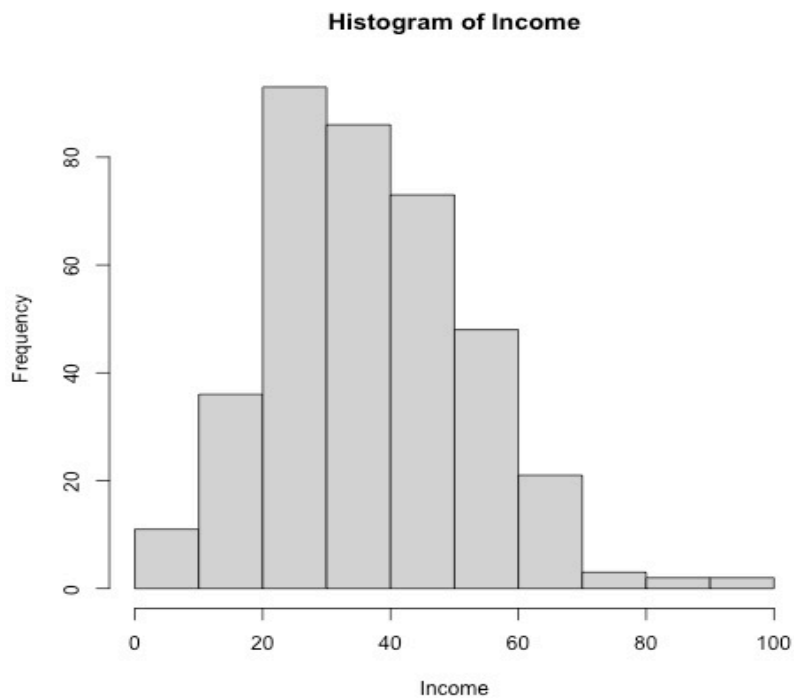
# The Credit Card Data with lavaan (Instrumental Variables)\*

```
> # Income and credit card debt
> # For a sample of real estate agents in different towns and cities, record
> # HomePrice: Median price of resale home, in 100k units
> # Income: Income last year after taxes, in thousands
> # CardDebt: Credit card balance carried forward, in thousands
>
> rm(list=ls())
> cards = read.table("https://www.utstat.toronto.edu/~brunner/openSEM/data/cards2.data.txt")
>
> # Explore the data
> head(cards); dim(cards)
  HomePrice Income CardDebt
1    7.918    45    1.122
2    7.765    40    2.711
3    6.533    54    0.000
4    4.908    27    2.539
5   10.235    64    4.759
6    7.089    35    2.740
[1] 375  3
> summary(cards)
  HomePrice      Income      CardDebt
Min.   : 1.724   Min.   : 0.00   Min.   :0.000
1st Qu.: 4.941   1st Qu.: 27.00   1st Qu.:0.355
Median : 5.935   Median : 37.00   Median :1.795
Mean   : 5.970   Mean   : 37.37   Mean   :2.092
3rd Qu.: 7.038   3rd Qu.: 47.00   3rd Qu.:3.471
Max.   :10.235   Max.   :100.00   Max.   :8.031
> cor(cards)
      HomePrice      Income      CardDebt
HomePrice 1.0000000 0.56473092 0.25016717
Income    0.5647309 1.00000000 0.01610061
CardDebt  0.2501672 0.01610061 1.00000000
> with(cards, hist(HomePrice) )
```

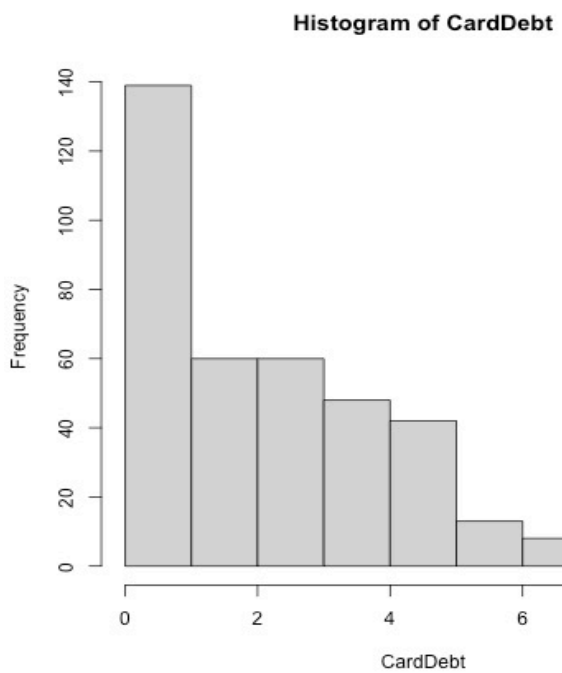


\* Copyright information is on the last page

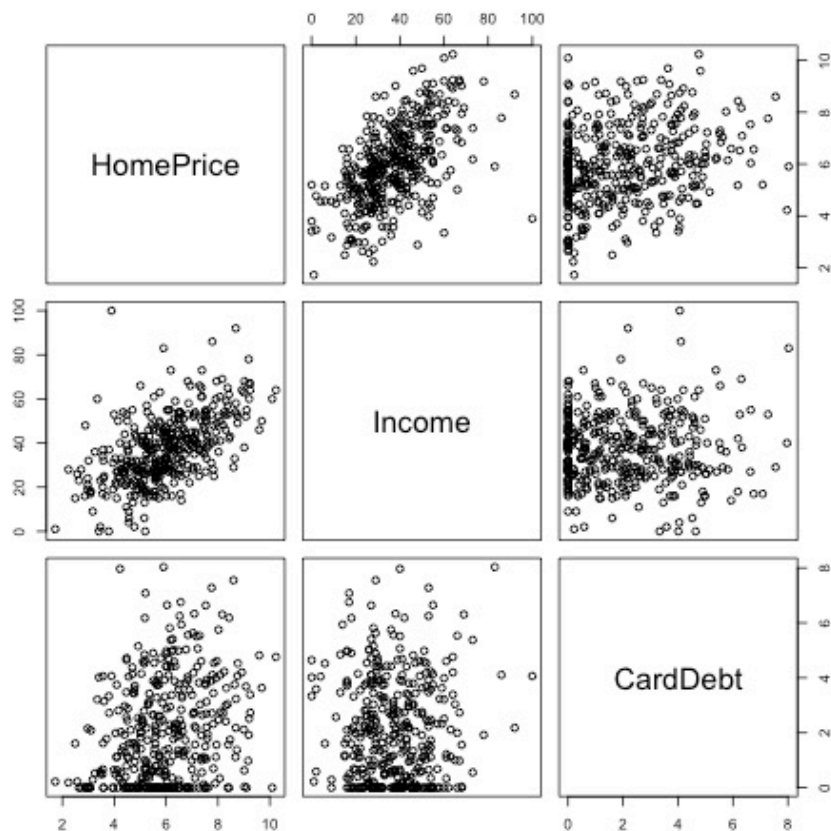
```
> with(cards, hist(Income) )
```



```
> with(cards, hist(CardDebt) )
```



```
> pairs(cards)
```



```
> # Ordinary regression  
> summary(lm(CardDebt ~ Income, data=cards))
```

```
Call:  
lm(formula = CardDebt ~ Income, data = cards)
```

```
Residuals:  
    Min      1Q  Median      3Q      Max  
-2.1510 -1.7399 -0.2724  1.3964  5.8702
```

```
Coefficients:  
              Estimate Std. Error t value Pr(>|t|)  
(Intercept)  2.019496   0.251897   8.017 1.41e-14 ***  
Income        0.001934   0.006218   0.311  0.756
```

```
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 1.883 on 373 degrees of freedom  
Multiple R-squared:  0.0002592,    Adjusted R-squared:  -0.002421  
F-statistic: 0.09672 on 1 and 373 DF,  p-value: 0.756
```

```

> # "Control" for median house price
> summary(lm(CardDebt ~ Income + HomePrice, data=cards))

Call:
lm(formula = CardDebt ~ Income + HomePrice, data = cards)

Residuals:
    Min       1Q   Median       3Q      Max
-3.362 -1.375 -0.289  1.239  6.975

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.354229   0.371586   0.953  0.34106
Income       -0.022074   0.007215  -3.059  0.00238 **
HomePrice    0.429259   0.072856   5.892 8.56e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.803 on 372 degrees of freedom
Multiple R-squared:  0.08559, Adjusted R-squared:  0.08067
F-statistic: 17.41 on 2 and 372 DF,  p-value: 5.919e-08

```

```

> # Now lavaan
> # install.packages("lavaan", dependencies = TRUE) # Only need to do this once
> library(lavaan)
This is lavaan 0.6-13
lavaan is FREE software! Please report any bugs.
>
> # Ordinary regression with random X
> mod0 = 'CardDebt ~ beta0*1 + beta1*Income
+       Income ~~ Phi*Income      # Var(Income) = phi
+       Income ~ mu*1             # E(Income) = mu
+       CardDebt ~~ psi*CardDebt  # Var(epsilon) = psi
+       ' # End of model string
>
> fit0 = lavaan(mod0, data=cards)
> summary(fit0)
lavaan 0.6.13 ended normally after 21 iterations

```

```

Estimator                               ML
Optimization method                     NLMINB
Number of model parameters              5

Number of observations                   375

```

Model Test User Model:

```

Test statistic                           0.000
Degrees of freedom                       0

```

Parameter Estimates:

```

Standard errors                          Standard
Information                               Expected
Information saturated (h1) model         Structured

```

Regressions:

	Estimate	Std.Err	z-value	P(> z )
CardDebt ~				
Income (bet1)	0.002	0.006	0.312	0.755

Intercepts:

	Estimate	Std.Err	z-value	P(> z )
.CardDbt (bet0)	2.019	0.251	8.039	0.000
Income (mu)	37.371	0.808	46.277	0.000

Variances:

	Estimate	Std.Err	z-value	P(> z )
Income (Phi)	244.548	17.859	13.693	0.000
.CardDebt (psi)	3.527	0.258	13.693	0.000

Repeating the last part of the lavaan output ...

```
Regressions:
      Estimate Std.Err z-value P(>|z|)
CardDebt ~
  Income (bet1)  0.002  0.006  0.312  0.755
```

```
Intercepts:
      Estimate Std.Err z-value P(>|z|)
.CardDbt (bet0)  2.019  0.251  8.039  0.000
Income (mu)  37.371  0.808  46.277  0.000
```

```
Variances:
      Estimate Std.Err z-value P(>|z|)
Income (Phi)  244.548  17.859  13.693  0.000
.CardDebt (psi)  3.527  0.258  13.693  0.000
```

```
> summary(lm(CardDebt ~ Income, data=cards))
```

```
Call:
lm(formula = CardDebt ~ Income, data = cards)
```

```
Residuals:
    Min     1Q   Median     3Q      Max
-2.1510 -1.7399 -0.2724  1.3964  5.8702
```

```
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.019496   0.251897   8.017 1.41e-14 ***
Income       0.001934   0.006218   0.311  0.756
---

```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 1.883 on 373 degrees of freedom
Multiple R-squared:  0.0002592,    Adjusted R-squared:  -0.002421
F-statistic: 0.09672 on 1 and 373 DF,  p-value: 0.756
```

```
>
> 1.883^2 # MSE from lm
[1] 3.545689
> c( mean(cards$Income), var(cards$Income) )
[1] 37.37067 245.20181
```

```

> # Instrumental variables model (Called Model Three in lecture)
> mod1 = 'CardDebt ~ beta0*1 + betal*Income
+       Income ~ mux*1           # E(Income)      = mux
+       HomePrice ~ muz*1        # E(HomePrice)   = muz
+       Income ~~ phix*Income    # Var(Income)   = phix
+       HomePrice ~~ phiz*HomePrice # Var(HomePrice) = phiz
+       Income ~~ kappa*HomePrice # Cov(Income,HomePrice) = kappa
+       CardDebt ~~ psi*CardDebt  # Var(epsilon)  = psi
+       Income ~~ c*CardDebt     # Cov(Income,epsilon) = c
+       ' # End of model string
> fit1 = lavaan(mod1, data=cards)
> summary(fit1)

```

lavaan 0.6.13 ended normally after 53 iterations

Estimator	ML
Optimization method	NLMINB
Number of model parameters	9
Number of observations	375

Model Test User Model:

Test statistic	0.000
Degrees of freedom	0

Parameter Estimates:

Standard errors	Standard
Information	Expected
Information saturated (h1) model	Structured

Regressions:

	Estimate	Std.Err	z-value	P(> z )
CardDebt ~				
Income (bet1)	0.053	0.012	4.456	0.000

Covariances:

	Estimate	Std.Err	z-value	P(> z )
Income ~~				
HomePrc (kapp)	13.677	1.436	9.522	0.000
.CardDebt ~~				
Income (c)	-12.538	2.495	-5.025	0.000

Intercepts:

	Estimate	Std.Err	z-value	P(> z )
.CardDbt (bet0)	0.103	0.459	0.226	0.821
Income (mux)	37.371	0.808	46.277	0.000
HomePrc (muz)	5.970	0.080	74.640	0.000

Variances:

	Estimate	Std.Err	z-value	P(> z )
Income (phix)	244.548	17.859	13.693	0.000
HomePrc (phiz)	2.399	0.175	13.693	0.000
.CardDbt (psi)	4.170	0.427	9.764	0.000

Repeating part of the lavaan output ...

```
Regressions:
      Estimate Std.Err z-value P(>|z|)
CardDebt ~
Income (bet1) 0.053 0.012 4.456 0.000
```

```
> # Checking explicit formula for the MLE.
> # (x,y,z) = (Income, CardDebt, HomePrice)
```

$$\hat{\beta}_1 = \hat{\Sigma}_{23} \hat{\Sigma}_{13}^{-1}$$

```
> with(cards, var(CardDebt,HomePrice) / var(Income,HomePrice) )
[1] 0.05320506
```

---

This document was prepared by [Jerry Brunner](#), University of Toronto. It is licensed under a Creative Commons Attribution - ShareAlike 3.0 Unported License:

[http://creativecommons.org/licenses/by-sa/3.0/deed.en\\_US](http://creativecommons.org/licenses/by-sa/3.0/deed.en_US). Use any part of it as you like and share the result freely. The Open Office document is available from the course website at <http://www.utstat.toronto.edu/~brunner/oldclass/431s23>