Introductory Stata Training
Lesson 3 – Analyzing Data

Harry Gimiseve, Rishabh Mukerjee, and Emily Schmidt

International Food Policy Research Institute

University of Papua New Guinea

March 2025

Port Moresby, PNG

Lesson 3 Agenda

1. Review of Lessons 1 and 2

Lesson 3 Agenda

1. Review of Lessons 1 and 2

2. Correlations

Lesson 3 Agenda

1. Review of Lessons 1 and 2
2. Correlations
3. T-Tests

Lesson 3 Agenda

1. Review of Lessons 1 and 2
2. Correlations
3. T-Tests
4. Ordinary least-squares (OLS) linear regressions

Lesson 3 Agenda

1. Review of Lessons 1 and 2
2. Correlations
3. T-Tests
4. Ordinary least-squares (OLS) linear regressions
5. Child Height-for-Age z-score (HAZ-score) presentation

Lesson 3 Agenda

1. Review of Lessons 1 and 2
2. Correlations
3. T-Tests
4. Ordinary least-squares (OLS) linear regressions
5. Child Height-for-Age z-score (HAZ-score) presentation
6. Child Height-for-Age z-score (HAZ-score) figures

1. Review of Lessons 1 and 2

- Today we are going to work a dataset that we compiled for this lesson, which is a compilation of key information from many different modules

1. Review of Lessons 1 and 2

- Today we are going to work a dataset that we compiled for this lesson, which is a compilation of key information from many different modules
- Let’s double-click to open *lesson3_data.dta*

1. Review of Lessons 1 and 2

- Today we are going to work a dataset that we compiled for this lesson, which is a compilation of key information from many different modules
- Let's double-click to open *lesson3.dta*
- How many observations are in this dataset?

1. Review of Lessons 1 and 2

- Today we are going to work a dataset that we compiled for this lesson, which is a compilation of key information from many different modules
- Let's double-click to open *lesson3.dta*
- How many observations are in this dataset? *Count – 1,334*

1. Review of Lessons 1 and 2

- Today we are going to work a dataset that we compiled for this lesson, which is a compilation of key information from many different modules
- Let's double-click to open *lesson3.dta*
- How many observations are in this dataset? *Count – 1,026*
- What level do you think this data is?

1. Review of Lessons 1 and 2

- Today we are going to work a dataset that we compiled for this lesson, which is a compilation of key information from many different modules
- Let’s double-click to open *lesson3.dta*
- How many observations are in this dataset? *Count – 1,026*
- What level do you think this data is? *Household-level*

1. Review of Lessons 1 and 2

- Let’s quickly take a look at some of the variables:

1. Review of Lessons 1 and 2

- Let's quickly take a look at some of the variables:
- How many households have a female household head?

1. Review of Lessons 1 and 2

- Let's quickly take a look at some of the variables:
- How many households have a female household head?
  - *tab hhh_female -*

1. Review of Lessons 1 and 2

- Let's quickly take a look at some of the variables:

- How many households have a female household head?
  - *tab hhh_female – 92 households (6.90%)*

1. Review of Lessons 1 and 2

- Let's quickly take a look at some of the variables:
- How many households have a female household head?
  - *tab hhh_female – 92 households (6.90%)*
- The roof variable has been cleaned a bit. What is the variable telling us now? What are the categories?

1. Review of Lessons 1 and 2

- Let's quickly take a look at some of the variables:
- How many households have a female household head?
  - *tab hhh_female – 92 households (6.90%)*
- The roof variable has been cleaned a bit. What is the variable telling us now? What are the categories?
  - *codebook roof –*

1. Review of Lessons 1 and 2

- Let's quickly take a look at some of the variables:
- How many households have a female household head?
  - *tab hhh_female – 92 households (6.90%)*
- The roof variable has been cleaned a bit. What is the variable telling us now? What are the categories?
  - *codebook roof – there are now 3 unique values or categories (thatched, metal, or other)*

1. Review of Lessons 1 and 2

- Let's quickly take a look at some of the variables:

- How about the new floor variable? What is the variable telling us now? What are the categories?
  - *codebook floor*

1. Review of Lessons 1 and 2

- Let’s quickly take a look at some of the variables:

- How about the new floor variable? What is the variable telling us now? What are the categories?
  - *codebook floor – tells us the type of material used to construct the floor of the main house); most individuals (568) have a floor made from Wild palm tree stem/strips, bamboo and pitipit (reed)/blind; followed by wood (494)*

1. Review of Lessons 1 and 2

- The HAZ score of children under 5 years is going to be our key outcome variable today.

1. Review of Lessons 1 and 2

- The Household Dietary Diversity Score (HDDS) will be our main variable of interest today
- . This variable was constructed in Module 8.1 in the survey – let’s look at that module – let’s all turn to this module in the questionnaire.

MODULE 8 – ANTHROPOMETRY

Section 8.1: Weight and height young child

We would like to take the weight and height measurement of all children under 5 years of age that are household members of this household. We would like to take their measurements to understand how your young children are growing. The measurement we will take today will not cause any harm to your children. Prior to measuring the children, we would like to measure the mother or caretaker of each child, this way, the children can see that the measurements will not cause harm. [ENUMERATOR: Please ask if the mother is present. If the mother is not present, ask if there is another who can be measured and then be responsible for holding the child (if under the age of 2) when the child is weighed].

Please ask the adult that will be measured to remove her shoes if possible.

children under 5years of age

Copy child information from Roster page *Children 5 years of age and under*											
1	2	3	4	5	6	7a	7b	8	9	10	11
Child Roster ID	Name	Sex	Age	Is there an adult (mother or other adult) that has agreed that the [CHILD] can be measured?	[ENUMERATOR]: Is the person who has agreed that [CHILD] can be measured a household member?	Who will assist [CHILD] for the weight measurement? (INSERT PID)	Is the [ADULT] that is assisting to measure the child the biological mother?	Are you currently pregnant, or do you think you may be pregnant?	Height of [**ADULT**] that is assisting in measurement?	Are you able to measure [CHILD]? Yes = 1 → *skip to Q17* No = 2	Why are you not able to measure [CHILD]? **[code: no_measr_vl]** → *Skip to next child*
		Male = 1 Female = 2		Yes=1 No=2 --> *skip to next child*	Yes=1 No=2 → *Skip to Q10*	Select from household roster If male → *Skip to Q10*	Yes = 1 No = 2	Yes = 1 No = 2 Don't know = 888 Refused = 999	In cm. to one decimal point	Yes = 1 No = 2	
									___ ___ ___ . ___		
									___ ___ ___ . ___		
									___ ___ ___ . ___		
									___ ___ ___ . ___		
									___ ___ ___ . ___		

Section 8.1 cont.: Weight and height of young child cont.

children under 5years of age

Copy child information from Roster page *Children 5 years of age and under*									
1	2	3	4	12	13	15	16	17	18
Child Roster ID	Name	Sex	Age	What is the weight of the **[ADULT]** helping to measure child? [ENUMERATOR: Please ask the mother / guardian to step onto the scale WITHOUT holding the baby / child.] IN KG TO ONE DECIMAL PLACE. (IF LESS THAN 10 KG, PUT ZERO IN FIRST BLANK.)	What is the weight of the **CHILD**? [**[ENUMERATOR: If child is under 2 years old**, press the 2 in 1 button in order **to tare the weight of the mother**, while the mother is still standing on the scale. **If child is over 2 years old**, please ask the child to step onto the scale]] IN KG TO ONE DECIMAL PLACE. (IF LESS THAN 10 KG, PUT ZERO IN FIRST BLANK.)	What is the height of the **CHILD**? [**ENUMERATOR: Children under 2 years** old should be measured lying down. **Children over 2 years** old and older should be measured standing.]] IN CM, TO ONE DECIMAL PLACE. (IF LESS THAN 100 CM, PUT ZERO IN FIRST BLANK.)	Was the child height/length measured with child lying down or standing up? 1= Standing 2= Lying down **[code: how_measr_vl]**	Was the measurement of the child done in a normal manner or was measurement difficult? 1 = Normal 2 = Difficult **[code: measr_manner_vl]**	Did the child appear to have edema (swelling that is not normal)? [This can be swelling of feet, face, and hands, often with sores or marks on the skin] Yes = 1 No = 2 Don't know = 888
		Male = 1 Female = 2							
				___ ___ . ___	___ ___ . ___	___ ___ ___ . ___			
				___ ___ . ___	___ ___ . ___	___ ___ ___ . ___			
				___ ___ . ___	___ ___ . ___	___ ___ ___ . ___			
				___ ___ . ___	___ ___ . ___	___ ___ ___ . ___			

Section 8.1 cont.: Weight and height of young child cont.

children under 5years of age

Copy child information from Roster page *Children 5 years of age and under*									
1	2	3	4	12	13	15	16	17	18
Child Roster ID	Name	Sex	Age	What is the weight of the **[ADULT]** helping to measure child? [ENUMERATOR: Please ask the mother / guardian to step onto the scale WITHOUT holding the baby / child.] IN KG TO ONE DECIMAL PLACE. (IF LESS THAN 10 KG, PUT ZERO IN FIRST BLANK.)	What is the weight of the **CHILD**? [**[ENUMERATOR: If child is under 2 years old**, press the 2 in 1 button in order **to tare the weight of the mother**, while the mother is still standing on the scale. **If child is over 2 years old**, please ask the child to step onto the scale]] IN KG TO ONE DECIMAL PLACE. (IF LESS THAN 10 KG, PUT ZERO IN FIRST BLANK.)	What is the height of the **CHILD**? [**ENUMERATOR**: **Children under 2 years** old should be measured lying down. **Children over 2 years** old and older should be measured standing.]] IN CM, TO ONE DECIMAL PLACE. (IF LESS THAN 100 CM, PUT ZERO IN FIRST BLANK.)	Was the child height/length measured with child lying down or standing up? 1= Standing 2= Lying down **[code: how_measr_vl]**	Was the measurement of the child done in a normal manner or was measurement difficult? 1 = Normal 2 = Difficult **[code: measr_manner_vl]**	Did the child appear to have edema (swelling that is not normal)? [This can be swelling of feet, face, and hands, often with sores or marks on the skin] Yes = 1 No = 2 Don't know = 888
		Male = 1 Female = 2							
				___ ___ . ___	___ ___ . ___	___ ___ ___ . ___			
				___ ___ . ___	___ ___ . ___	___ ___ ___ . ___			
				___ ___ . ___	___ ___ . ___	___ ___ ___ . ___			
				___ ___ . ___	___ ___ . ___	___ ___ ___ . ___			

Weight and Height of Child under 5 years

1. Review of Lessons 1 and 2

- To better understand nutrition and health outcomes of children in the survey sample, the 2023 PNG Rural Household Survey collected anthropometric measurements (height and weight) for all children under 5 years old in the survey sample. In doing so, the survey data allow for estimation of under-5 child statistics on undernutrition including stunting, wasting, and underweight indicators.
- To remove biologically impossible scores (extreme outliers), a child haz-score greater than 6 or less than -6 was changed to missing, in accordance with World Health Organization's (WHO) standards.

1. Review of Lessons 1 and 2

- To better understand nutrition and health outcomes of children in the survey sample, the 2023 PNG Rural Household Survey collected anthropometric measurements (height and weight) for all children under 5 years old in the survey sample. In doing so, the survey data allow for estimation of under-5 child statistics on undernutrition including stunting, wasting, and underweight indicators.
- To remove biologically impossible scores (extreme outliers), a child haz-score greater than 6 or less than -6 was changed to missing, in accordance with World Health Organization's (WHO) standards.
- A higher haz-score indicates healthy growth outcomes

1. Review of Lessons 1 and 2

- To better understand nutrition and health outcomes of children in the survey sample, the 2023 PNG Rural Household Survey collected anthropometric measurements (height and weight) for all children under 5 years old in the survey sample. In doing so, the survey data allow for estimation of under-5 child statistics on undernutrition including stunting, wasting, and underweight indicators.
- To remove biologically impossible scores (extreme outliers), a child haz-score greater than 6 or less than -6 was changed to missing, in accordance with World Health Organization's (WHO) standards.
- A higher haz-score indicates healthy growth outcomes
- A higher score indicates higher household dietary diversity
- What do you think the minimum and maximum values are for HDDS?

1. Review of Lessons 1 and 2

- To better understand nutrition and health outcomes of children in the survey sample, the 2023 PNG Rural Household Survey collected anthropometric measurements (height and weight) for all children under 5 years old in the survey sample. In doing so, the survey data allow for estimation of under-5 child statistics on undernutrition including stunting, wasting, and underweight indicators.
- To remove biologically impossible scores (extreme outliers), a child haz-score greater than 6 or less than -6 was changed to missing, in accordance with World Health Organization's (WHO) standards.
- A higher haz-score indicates healthy growth outcomes
- What do you think the minimum and maximum values are for the variable haz? *-5.94 and 5.9*

1. Review of Lessons 1 and 2

- To better understand nutrition and health outcomes of children in the survey sample, the 2023 PNG Rural Household Survey collected anthropometric measurements (height and weight) for all children under 5 years old in the survey sample. In doing so, the survey data allow for estimation of under-5 child statistics on undernutrition including stunting, wasting, and underweight indicators.
- To remove biologically impossible scores (extreme outliers), a child haz-score greater than 6 or less than -6 was changed to missing, in accordance with World Health Organization's (WHO) standards.
- A higher haz-score indicates healthy growth outcomes
- What do you think the minimum and maximum values are for the variable haz? -5.94 and 5.9
- How can we check?

1. Review of Lessons 1 and 2

- To better understand nutrition and health outcomes of children in the survey sample, the 2023 PNG Rural Household Survey collected anthropometric measurements (height and weight) for all children under 5 years old in the survey sample. In doing so, the survey data allow for estimation of under-5 child statistics on undernutrition including stunting, wasting, and underweight indicators.
- To remove biologically impossible scores (extreme outliers), a child haz-score greater than 6 or less than -6 was changed to missing, in accordance with World Health Organization's (WHO) standards.
- A higher haz-score indicates healthy growth outcomes
- What do you think the minimum and maximum values are for the variable haz? -5.94 and 5.9
- How can we check? *codebook haz* OR *sum haz*

1. Review of Lessons 1 and 2

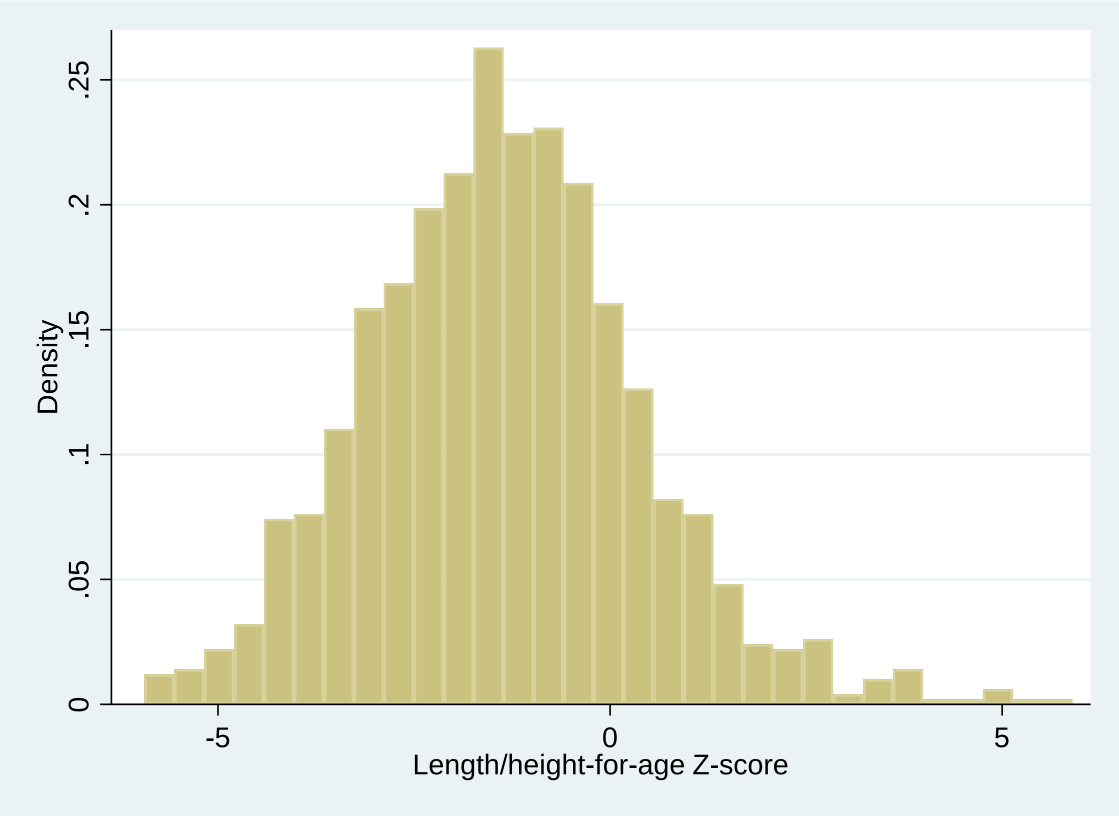
- Let’s see what the distribution of haz looks like:

1. Review of Lessons 1 and 2

- Let’s see what the distribution of haz looks like:
- *hist haz*

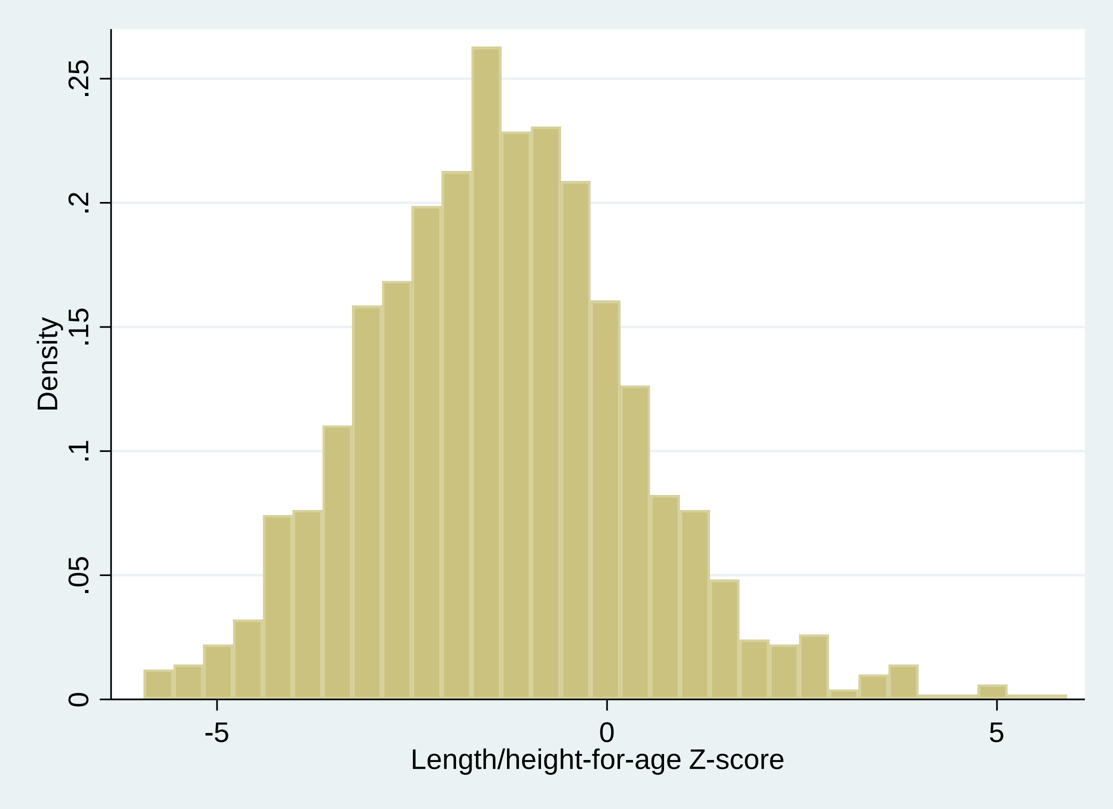
1. Review of Lessons 1 and 2

- Let's see what the distribution of haz looks like:
- *hist haz*



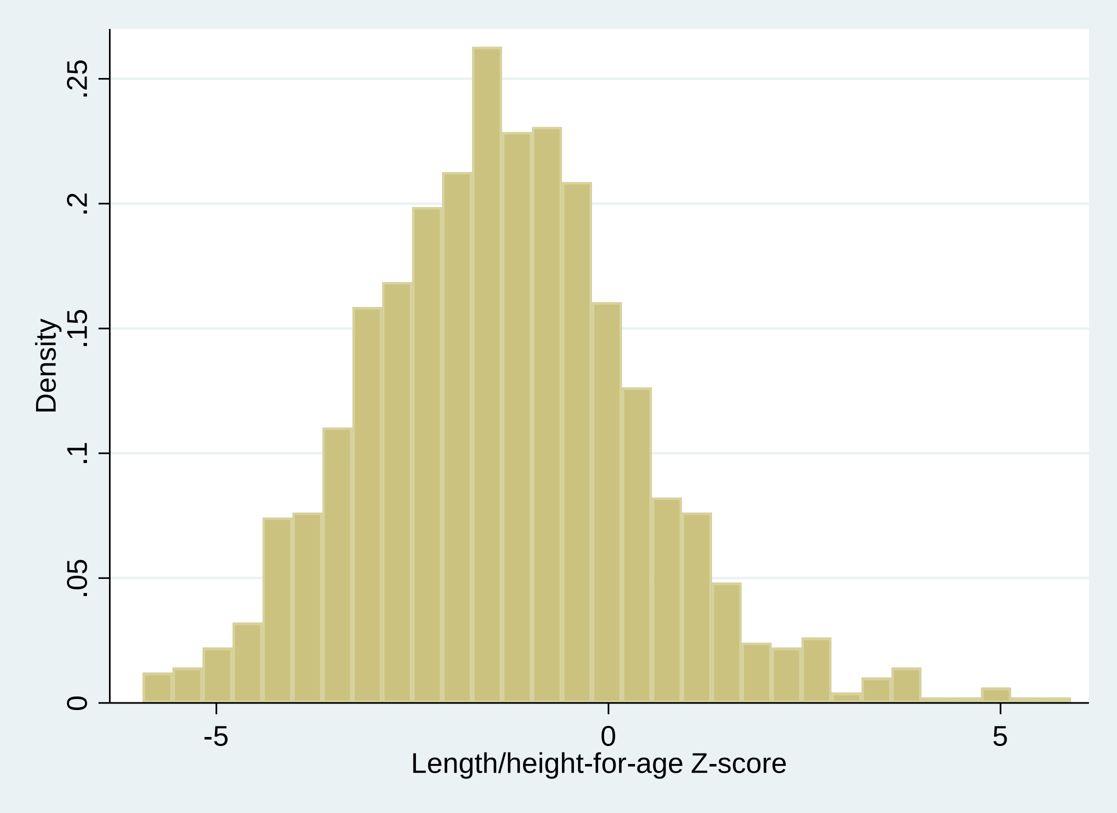
1. Review of Lessons 1 and 2

- Let's see what the distribution of haz looks like:
- *hist haz*
- What can we learn from this figure?



1. Review of Lessons 1 and 2

- Let's see what the distribution of haz looks like:
  - *hist haz*
  - What can we learn from this figure? *Majority of children aged 5 or less in the surveyed areas have haz-scores that falls below 0 (WHO growth standard). A negative haz-score indicates that a child is shorter than WHO growth standard. (We have a bell-shaped curve)*



1. Review of Lessons 1 and 2

- We have some continuous data in this dataset
  - Examples: landholdings and household size (hhsize)

1. Review of Lessons 1 and 2

- We have some continuous data in this dataset
  - Examples: landholdings and household size (hhsize)

- Household size (hhsize) variable was generated by us, but landholdings is the raw data (directly reported by respondents)

1. Review of Lessons 1 and 2

- We have some continuous data in this dataset
  - Examples: landholdings and household size (hhsize)
- Household size (hhsize) variable was generated by us, but landholdings is the raw data (directly reported by respondents)
- Frequently, raw continuous data can be messy. For example:
  - Input errors
  - Question wording confusion
  - Best guesses

1. Review of Lessons 1 and 2

- Many times there are notable *outliers* in raw, continuous data
  - Outlier – an observation very different from all other observations

1. Review of Lessons 1 and 2

- Many times there are notable *outliers* in raw, continuous data
  - Outlier – an observation very different from all other observations
- What's a good way to check landholdings for outliers?

1. Review of Lessons 1 and 2

- Many times there are notable *outliers* in raw, continuous data
  - Outlier – an observation very different from all other observations
- What's a good way to check landholdings for outliers?
  - *sum landholdings, det*

1. Review of Lessons 1 and 2

- Many times there are notable *outliers* in raw, continuous data
  - Outlier – an observation very different from all other observations
- What's a good way to check landholdings for outliers?
  - *sum landholdings, det*

```
. sum landholdings, det

                Size of all landholdings in hectares
-------------------------------------------------------------
      Percentiles      Smallest
 1%            0              0
 5%        .0162              0
10%        .0324              0       Obs               1,334
25%        .3743              0       Sum of Wgt.       1,334

50%         1.05                      Mean           1.455915
                        Largest       Std. Dev.      1.688366
75%            2           10.7
90%          3.5           11.9       Variance       2.850578
95%         4.55           14.4       Skewness       3.122489
99%       8.0824          19.55       Kurtosis        20.5761
```

1. Review of Lessons 1 and 2

- Many times there are notable *outliers* in raw, continuous data
  - Outlier – an observation very different from all other observations
- What's a good way to check landholdings for outliers?
  - *sum landholdings, det*

```
. sum landholdings, det

                 Size of all landholdings in hectares
-------------------------------------------------------------
      Percentiles      Smallest
 1%            0              0
 5%        .0162              0
10%        .0324              0       Obs               1,334
25%        .3743              0       Sum of Wgt.       1,334

50%         1.05                      Mean           1.455915
                        Largest       Std. Dev.      1.688366
75%            2           10.7
90%          3.5           11.9       Variance       2.850578
95%         4.55           14.4       Skewness       3.122489
99%       8.0824          19.55       Kurtosis        20.5761
```

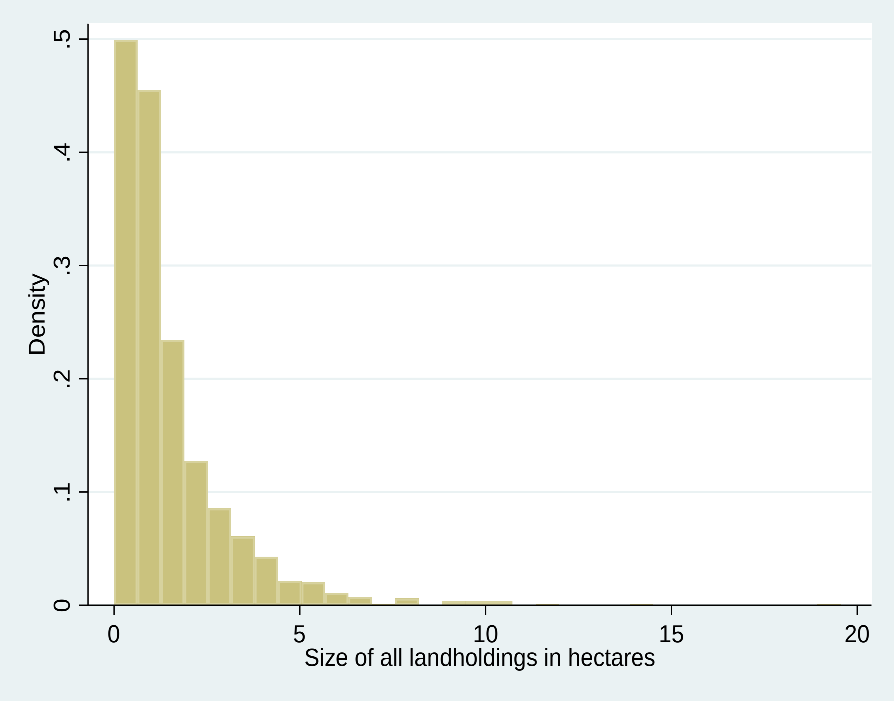
The median is 1.05, but the max is 19.55!

1. Review of Lessons 1 and 2

- Many times there are notable *outliers* in raw, continuous data
  - Outlier – an observation very different from all other observations
- What's a good way to check landholdings for outliers?
  - *sum landholdings, det*
  - *hist landholdings*

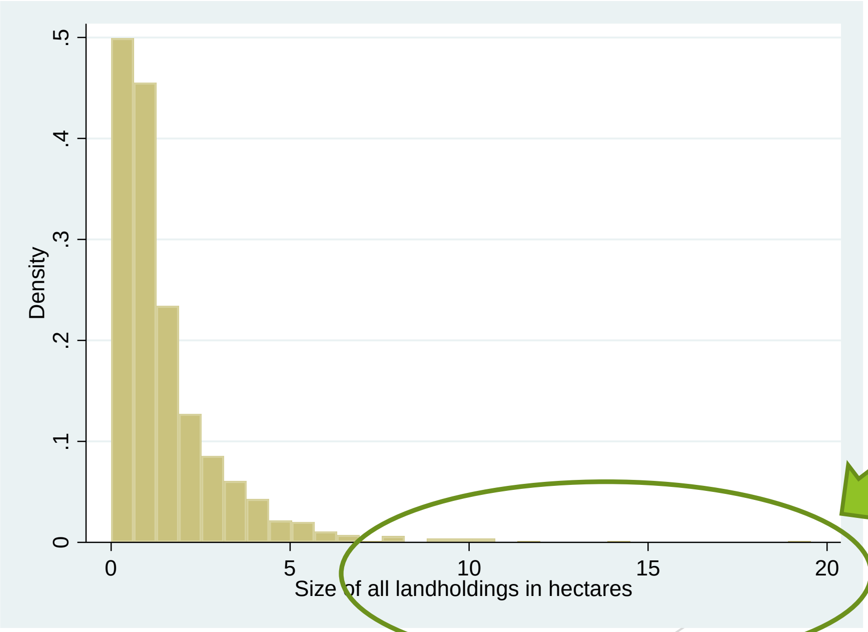
1. Review of Lessons 1 and 2

- Many times there are notable *outliers* in raw, continuous data
  - Outlier – an observation very different from all other observations
- What's a good way to check landholdings for outliers?
  - *sum landholdings, det*
  - *hist landholdings*



1. Review of Lessons 1 and 2

- Many times there are notable *outliers* in raw, continuous data
  - Outlier – an observation very different from all other observations
- What's a good way to check landholdings for outliers?
  - *sum landholdings, det*
  - *hist landholdings*



Very positively skewed (a few, very large obs. on the righthand side)

1. Review of Lessons 1 and 2

- Everyone has a different preference for handling outliers

1. Review of Lessons 1 and 2

- Everyone has a different preference for handling outliers
- Let's change all values greater than the 99^{th} percentile to the median

1. Review of Lessons 1 and 2

- Everyone has a different preference for handling outliers
- Let's change all values greater than the 99th percentile to the median
- First, let's check the median and the 99th percentile again
  - *sum landholdings, det*

1. Review of Lessons 1 and 2

- Everyone has a different preference for handling outliers
- Let's change all values greater than the 99th percentile to the median
- First, let's check the median and the 99th percentile again
  - *sum landholdings, det*

```
. sum landholdings, det

              Size of all landholdings in hectares
-------------------------------------------------------------
      Percentiles      Smallest
 1%            0              0
 5%        .0162              0
10%        .0324              0       Obs               1,334
25%        .3743              0       Sum of Wgt.       1,334

50%         1.05                      Mean           1.455915
                        Largest       Std. Dev.      1.688366
75%            2           10.7
90%          3.5           11.9       Variance       2.850578
95%         4.55           14.4       Skewness       3.122489
99%       8.0824          19.55       Kurtosis        20.5761
```

1. Review of Lessons 1 and 2

- Everyone has a different preference for handling outliers
- Let's change all values greater than the 99th percentile to the median
- First, let's check the median and the 99th percentile again
  - *sum landholdings, det*
  - *Median: 1.05*

```
. sum landholdings, det

              Size of all landholdings in hectares
-------------------------------------------------------------
      Percentiles      Smallest
 1%            0              0
 5%        .0162              0
10%        .0324              0       Obs               1,334
25%        .3743              0       Sum of Wgt.       1,334

50%         1.05                      Mean           1.455915
                        Largest       Std. Dev.      1.688366
75%            2           10.7
90%          3.5           11.9       Variance       2.850578
95%         4.55           14.4       Skewness       3.122489
99%       8.0824          19.55       Kurtosis        20.5761
```

1. Review of Lessons 1 and 2

- Everyone has a different preference for handling outliers
- Let's change all values greater than the 99th percentile to the median
- First, let's check the median and the 99th percentile again
  - *sum landholdings, det*
  - *Median: 1.05*
  - *99th percentile: 8.0824*

```
. sum landholdings, det

              Size of all landholdings in hectares
-------------------------------------------------------------
      Percentiles      Smallest
 1%            0              0
 5%        .0162              0
10%        .0324              0       Obs               1,334
25%        .3743              0       Sum of Wgt.       1,334

50%         1.05                      Mean           1.455915
                        Largest       Std. Dev.      1.688366
75%            2           10.7
90%          3.5           11.9       Variance       2.850578
95%         4.55           14.4       Skewness       3.122489
99%       8.0824          19.55       Kurtosis        20.5761
```

1. Review of Lessons 1 and 2

- How can we check how many observations are above the 99th percentile (8.0824 hectares)?
  - **HINT:** it combines a code we use every time we open a new dataset, with "if"

1. Review of Lessons 1 and 2

- How can we check how many observations are above the 99th percentile (8.0824 hectares)?
  - **HINT:** it combines a code we use every time we open a new dataset, with "if"
  - *count if landholdings>8.0824*

1. Review of Lessons 1 and 2

- How can we check how many observations are above the 99th percentile (8.0824 hectares)?
  - **HINT:** it combines a code we use every time we open a new dataset, with "if"
  - *count if landholdings>8.0824*
  - *12 observations in landholdings are greater than the 99th percentile (8.0824 hectares)*

```
. count if landholdings>8.0824
  12
```

1. Review of Lessons 1 and 2

- How can we change these observations to the median (1.05 hectares)?

1. Review of Lessons 1 and 2

- How can we change these observations to the median (1.05 hectares)?
  - *Replace landholding=1.05 if landholdings>8.0824*

1. Review of Lessons 1 and 2

- How can we change these observations to the median (1.05 hectares)?
  - *Replace landholding=1.05 if landholdings>8.0824*
- How many observations were changed?

1. Review of Lessons 1 and 2

- How can we change these observations to the median (1.05 hectares)?
  - *Replace landholding=1.05 if landholdings>8.0824*

- How many observations were changed? *12 observations*

```
. replace landholdings=1.05 if landholdings>8.0824
(12 real changes made)
```

1. Review of Lessons 1 and 2

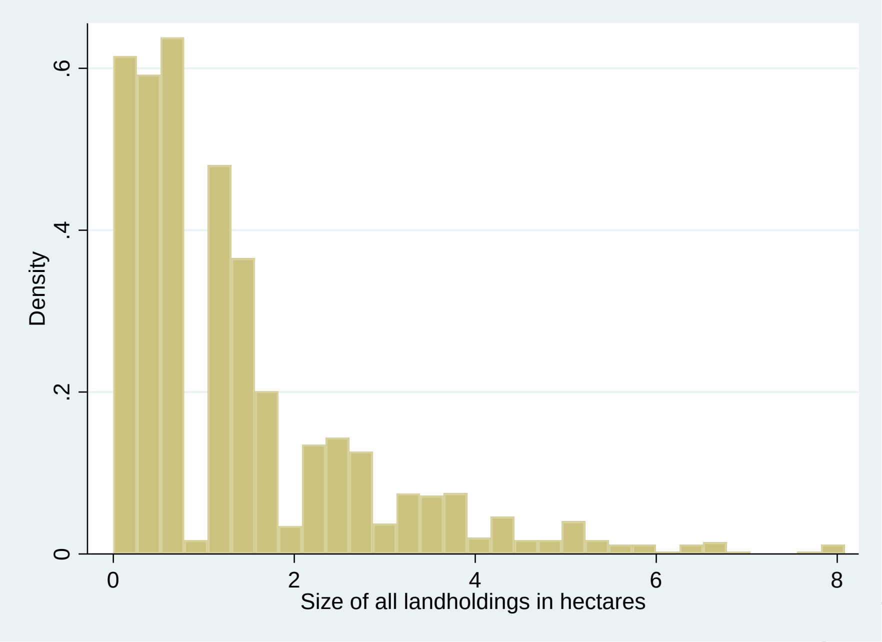
- Now let’s see what the new distribution looks like

1. Review of Lessons 1 and 2

- Now let's see what the new distribution looks like
- *hist landholdings*

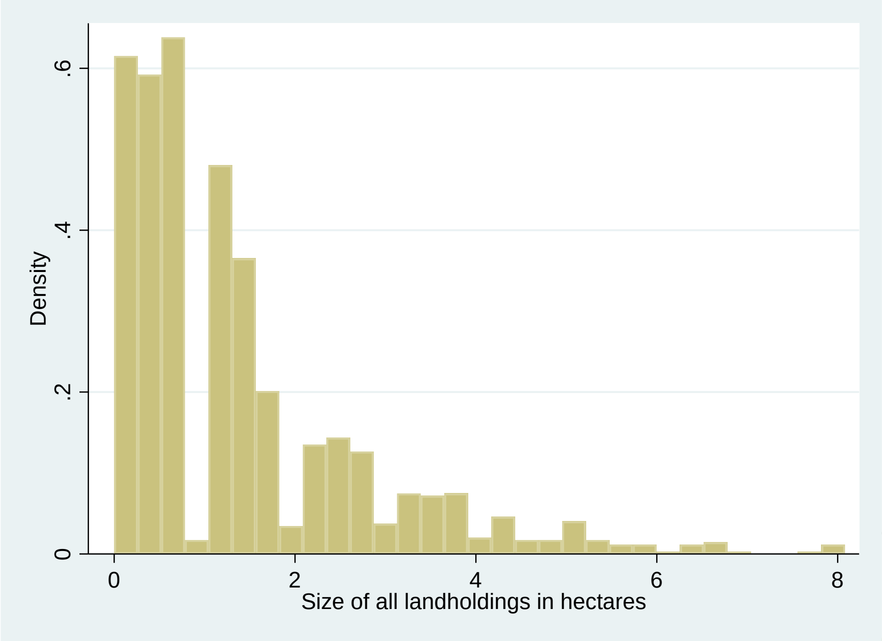
1. Review of Lessons 1 and 2

- Now let's see what the new distribution looks like
- *hist landholdings*



1. Review of Lessons 1 and 2

- Now let's see what the new distribution looks like
- *hist landholdings*
- Still positively skewed, but now the figure only goes up to 8 hectares – before it was over 20!



2. Correlations

- Let's see how correlated our variable of interest (haz) is with household real daily consumption expenditure variable (ex_r).

2. Correlations

- Let’s see how correlated our variable of interest (haz) is with household real daily consumption expenditure variable (ex_r).
- Code: *pwcorr haz ex_r*

2. Correlations

- Let’s see how correlated our variable of interest (haz) is with household real daily consumption expenditure variable (ex_r).
- Code: *pwcorr haz ex_r*
  - This only tells us the correlation coefficient between the two variables

2. Correlations

- Let's see how correlated our variable of interest (haz) is with household real daily consumption expenditure variable (ex_r).
- Code: *pwcorr haz ex_r*
  - This only tells us the correlation coefficient between the two variables
  - Remember that a correlation coefficient equal to 0 is the weakest linear relationship, and a correlation coefficient equal to 1 or -1 is the strongest linear relationship

2. Correlations

- Let's see how correlated our variable of interest (haz) is with household real daily consumption expenditure variable (ex_r).
- Code: *pwcorr haz ex_r*
  - This only tells us the correlation coefficient between the two variables
  - Remember that a correlation coefficient equal to 0 is the weakest linear relationship, and a correlation coefficient equal to 1 or -1 is the strongest linear relationship
  - A positive coefficient means that as one variable increases, the other increases, whereas a negative coefficient means that as one variable increases, the other decreases

2. Correlations

- *pwcorr haz ex_r*

```
. pwcorr haz ex_r

             |      haz     ex_r
-------------+------------------
         haz |   1.0000
        ex_r |   0.1194   1.0000
```

2. Correlations

- *pwcorr haz ex_r*
- What is the correlation coefficient?

```
. pwcorr haz ex_r

             |      haz     ex_r
-------------+------------------
         haz |   1.0000
        ex_r |   0.1194   1.0000
```

2. Correlations

- *pwcorr haz ex_r*

- What is the correlation coefficient? *0.1194*

```
. pwcorr haz ex_r

             |      haz     ex_r
-------------+------------------
         haz |   1.0000
        ex_r |   0.1194   1.0000
```

2. Correlations

- *pwcorr haz ex_r*
- What is the correlation coefficient? *0.1194*
- Is this a strong or weak coefficient?

```
. pwcorr haz ex_r

             |      haz     ex_r
-------------+------------------
         haz |   1.0000
        ex_r |   0.1194   1.0000
```

2. Correlations

- *pwcorr haz ex_r*

- What is the correlation coefficient? *0.1194*
- Is this a strong or weak coefficient? *Weak*

```
. pwcorr haz ex_r

             |      haz     ex_r
-------------+------------------
         haz |   1.0000
        ex_r |   0.1194   1.0000
```

2. Correlations

- *pwcorr haz ex_r*
- What is the correlation coefficient? *0.1194*
- *Is this a strong or weak coefficient? Weak*
- Is this a positive or negative coefficient?

```
. pwcorr haz ex_r

             |      haz     ex_r
-------------+------------------
         haz |   1.0000
        ex_r |   0.1194   1.0000
```

2. Correlations

- *pwcorr haz ex_r*
- What is the correlation coefficient? *0.1194*
- *Is this a strong or weak coefficient? Weak*
- Is this a positive or negative coefficient? *Positive*

```
. pwcorr haz ex_r

             |      haz     ex_r
-------------+------------------
         haz |   1.0000 
        ex_r |   0.1194   1.0000 
```

2. Correlations

- Even though this coefficient is very weak, how can we interpret this?

```
. pwcorr haz ex_r

             |      haz     ex_r
-------------+------------------
         haz |   1.0000
        ex_r |   0.1194   1.0000
```

2. Correlations

- Even though this coefficient is very weak, how can we interpret this?
  - *Households with higher consumption expenditure are associated with children being less likely to be too short for their age.*

```
. pwcorr haz ex_r

             |      haz     ex_r
-------------+------------------
         haz |   1.0000
        ex_r |   0.1194   1.0000
```

2. Correlations

- We can also look at how significant a correlation coefficient is, by adding the “sig” option to the pwcorr code

2. Correlations

- We can also look at how significant a correlation coefficient is, by adding the “sig” option to the pwcorr code

- Code: *pwcorr haz ex_r, sig*

2. Correlations

- We can also look at how significant a correlation coefficient is, by adding the "sig" option to the pwcorr code

- Code: *pwcorr haz ex_r, sig*

```
. pwcorr haz ex_r, sig

             |      haz     ex_r
-------------+------------------
         haz |   1.0000
             |
             |
        ex_r |   0.1194   1.0000
             |   0.0000
             |
```

2. Correlations

- We can also look at how significant a correlation coefficient is, by adding the "sig" option to the pwcorr code

- Code: *pwcorr haz ex_r, sig*

- What is the p-value?

```
. pwcorr haz ex_r, sig

             |      haz     ex_r
-------------+------------------
         haz |   1.0000
             |
             |
        ex_r |   0.1194   1.0000
             |   0.0000
             |
```

2. Correlations

- We can also look at how significant a correlation coefficient is, by adding the “sig” option to the pwcorr code

- Code: *pwcorr haz ex_r, sig*

- What is the p-value? *0.0000*

```
. pwcorr haz ex_r, sig

             |      haz     ex_r
-------------+------------------
         haz |   1.0000
             |
             |
        ex_r |   0.1194   1.0000
             |   0.0000
             |
```

2. Correlations

- We can also look at how significant a correlation coefficient is, by adding the “sig” option to the pwcorr code

- Code: *pwcorr haz ex_r, sig*

- What is the p-value? *0.0000*
- Is this significant?

```
. pwcorr haz ex_r, sig

             |      haz     ex_r
-------------+------------------
         haz |   1.0000 
             |
             |
        ex_r |   0.1194   1.0000 
             |   0.0000
             |
```

2. Correlations

- We can also look at how significant a correlation coefficient is, by adding the "sig" option to the pwcorr code
- Code: *pwcorr haz ex_r, sig*
- What is the p-value? *0.0000*
- Is this significant? *Yes*

```
. pwcorr haz ex_r, sig

             |      haz     ex_r
-------------+------------------
         haz |   1.0000
             |
             |
        ex_r |   0.1194   1.0000
             |   0.0000
             |
```

2. Correlations

- This p-value is saying that we can only be 100% confident (100-0.0000) that the correlation coefficient between haz and ex_r is significant

```
. pwcorr haz ex_r, sig

             |      haz     ex_r
-------------+------------------
         haz |   1.0000
             |
             |
        ex_r |   0.1194   1.0000
             |   0.0000
             |
```

2. Correlations

- This p-value is saying that we can only be 100% confident (100-0.0000) that the correlation coefficient between haz and ex_r is significant

```
. pwcorr haz ex_r, sig

             |      haz     ex_r
-------------+------------------
         haz |   1.0000
             |
             |
        ex_r |   0.1194   1.0000
             |   0.0000
             |
```

- Usual significance cut-offs are
  - 90% confidence (p-value=0.1)
  - 95% confidence (p-value=0.05)
  - 99% confidence (p-value=0.01)

2. Correlations

- Another way to look at this relationship between haz and ex_r is to create a scatterplot

2. Correlations

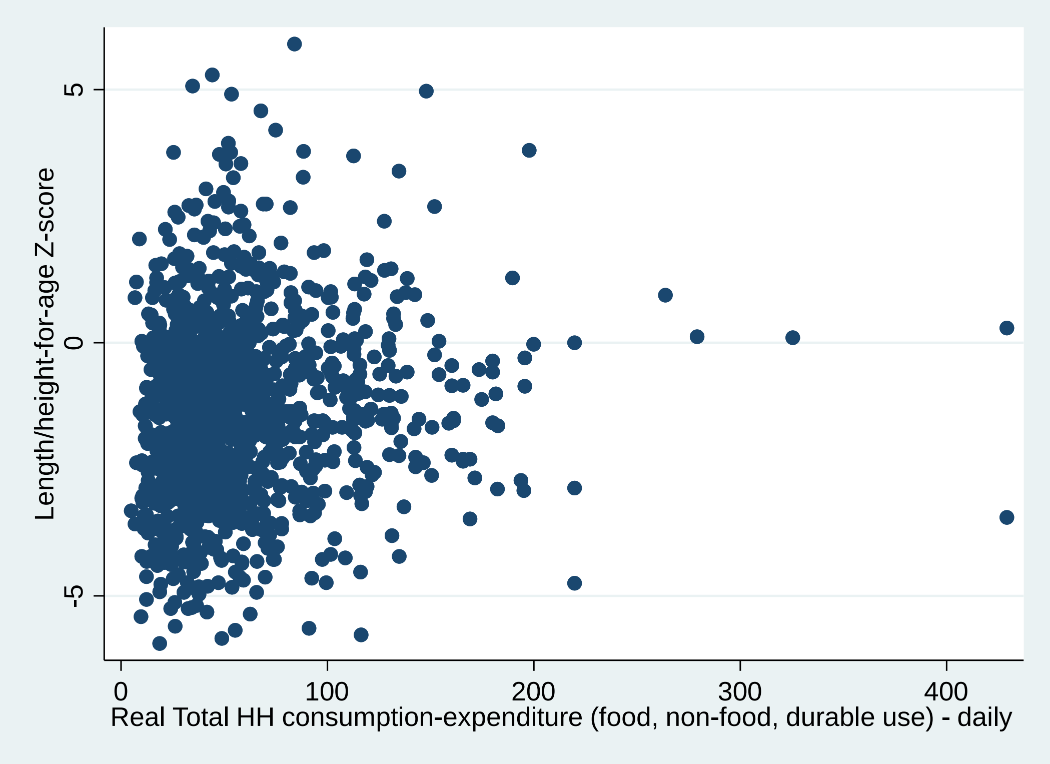
- Another way to look at this relationship between haz and ex_r is to create a scatterplot

- Code: *scatter haz ex_r*

2. Correlations

- Another way to look at this relationship between haz and ex_r is to create a scatterplot

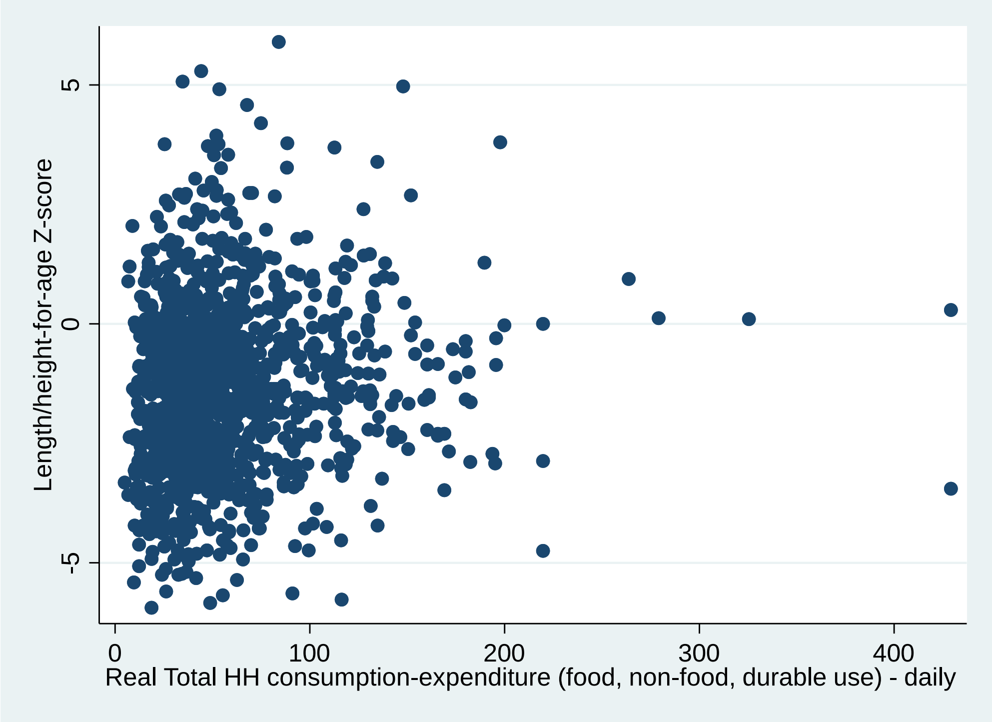
- Code: *scatter haz ex_r*



2. Correlations

- Another way to look at this relationship between haz and ex_r is to create a scatterplot
- Code: *scatter haz ex_r*

- It looks like there may be a positive relationship, but we can also create a 'line of best fit' to see how positive and strong the relationship is

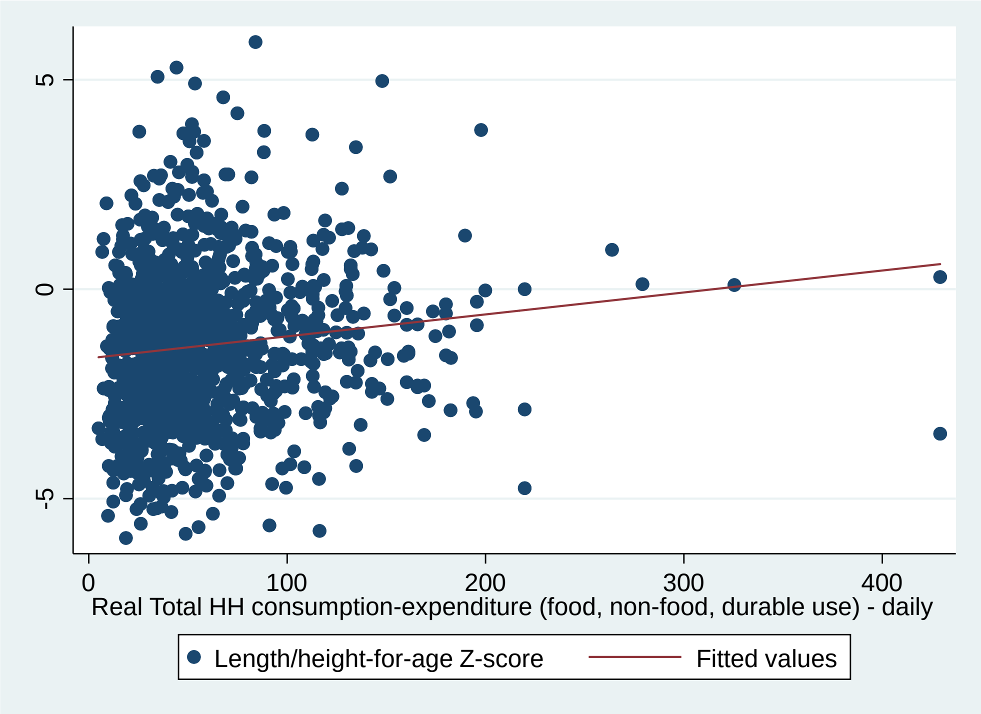


2. Correlations

- Code: *twoway (scatter haz ex_r) (lfit haz ex_r)*

2. Correlations

- Code: twoway (scatter haz ex_r) (lfit haz ex_r)



3. T-Tests

- The most frequently used t-tests are two-sample t-tests

3. T-Tests

- The most frequently used t-tests are two-sample t-tests
- These tell us whether one variable (e.g. haz) is significantly different between two groups in the data (e.g. whether or not a chlild under 5 years eat meat, eggs or poultry in the last 24 hours)

3. T-Tests

- Code: *ttest haz, by(meat_poul_fish_C)*

3. T-Tests

- Code: *ttest haz, by(meat_poul_fish_C)*

```
. ttest haz, by( meat_poul_fish_C)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |     224   -1.950223    .0990268    1.482097   -2.145371   -1.755075
     Yes |     374   -1.637834    .0827858    1.601001    -1.80062   -1.475049
---------+--------------------------------------------------------------------
combined |     598   -1.754849    .0639408     1.56361   -1.880426   -1.629273
---------+--------------------------------------------------------------------
    diff |            -.312389    .1315951               -.5708356   -.0539424
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t =  -2.3739
Ho: diff = 0                                     degrees of freedom =      596

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0090         Pr(|T| > |t|) = 0.0179          Pr(T > t) = 0.9910
```

3. T-Tests

- Code: ttest haz, by(meat_poul_fish_C)

```
. ttest haz, by( meat_poul_fish_C)

Two-sample t test with equal variances

   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |     224   -1.950223    .0990268    1.482097   -2.145371   -1.755075
     Yes |     374   -1.637834    .0827858    1.601001    -1.80062   -1.475049
---------+--------------------------------------------------------------------
combined |     598   -1.754849    .0639408     1.56361   -1.880426   -1.629273
---------+--------------------------------------------------------------------
    diff |            -.312389    .1315951               -.5708356   -.0539424
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t =  -2.3739
Ho: diff = 0                                     degrees of freedom =      596

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0090         Pr(|T| > |t|) = 0.0179          Pr(T > t) = 0.9910
```

- What does this output tell us?

3. T-Tests

- Code: ttest haz, by(meat_poul_fish_C)

```
. ttest haz, by( meat_poul_fish_C)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |     224   -1.950223    .0990268    1.482097   -2.145371   -1.755075
     Yes |     374   -1.637834    .0827858    1.601001    -1.80062   -1.475049
---------+--------------------------------------------------------------------
combined |     598   -1.754849    .0639408     1.56361   -1.880426   -1.629273
---------+--------------------------------------------------------------------
    diff |             -.312389    .1315951               -.5708356   -.0539424
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t =  -2.3739
Ho: diff = 0                                     degrees of freedom =      596

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0090         Pr(|T| > |t|) = 0.0179          Pr(T > t) = 0.9910
```

- The number of observations in each group

3. T-Tests

- Code: ttest haz, by(meat_poul_fish_C)

```
. ttest haz, by( meat_poul_fish_C)

Two-sample t test with equal variances

   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |     224   -1.950223    .0990268    1.482097   -2.145371   -1.755075
     Yes |     374   -1.637834    .0827858    1.601001    -1.80062   -1.475049
---------+--------------------------------------------------------------------
combined |     598   -1.754849    .0639408     1.56361   -1.880426   -1.629273
---------+--------------------------------------------------------------------
    diff |            -.312389    .1315951               -.5708356   -.0539424
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t =  -2.3739
Ho: diff = 0                                     degrees of freedom =      596

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0090         Pr(|T| > |t|) = 0.0179          Pr(T > t) = 0.9910
```

- How many observations do not eat meat, poultry or fish?
  - *224 children aged 2 to 5 years do not eat meat, poultry or fish in the last 24 hours.*

3. T-Tests

- Code: ttest haz, by(meat_poul_fish_C)

```
. ttest haz, by( meat_poul_fish_C)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |     224   -1.950223    .0990268    1.482097   -2.145371   -1.755075
     Yes |     374   -1.637834    .0827858    1.601001    -1.80062   -1.475049
---------+--------------------------------------------------------------------
combined |     598   -1.754849    .0639408     1.56361   -1.880426   -1.629273
---------+--------------------------------------------------------------------
    diff |            -.312389    .1315951               -.5708356   -.0539424
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t =  -2.3739
Ho: diff = 0                                     degrees of freedom =      596

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0090         Pr(|T| > |t|) = 0.0179          Pr(T > t) = 0.9910
```

- The average haz for each group

3. T-Tests

- Code: ttest haz, by(meat_poul_fish_C)

```
. ttest haz, by( meat_poul_fish_C)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |     224   -1.950223    .0990268    1.482097   -2.145371   -1.755075
     Yes |     374   -1.637834    .0827858    1.601001    -1.80062   -1.475049
---------+--------------------------------------------------------------------
combined |     598   -1.754849    .0639408     1.56361   -1.880426   -1.629273
---------+--------------------------------------------------------------------
    diff |            -.312389    .1315951               -.5708356   -.0539424
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t =  -2.3739
Ho: diff = 0                                     degrees of freedom =      596

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0090         Pr(|T| > |t|) = 0.0179          Pr(T > t) = 0.9910
```

- Which group has a higher average haz-score?

3. T-Tests

- Code: ttest haz, by(meat_poul_fish_C)

. **ttest haz, by(meat_poul_fish_C)**

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
No	224	-1.950223	.0990268	1.482097	-2.145371	-1.755075
Yes	374	-1.637834	.0827858	1.601001	-1.80062	-1.475049
combined	598	-1.754849	.0639408	1.56361	-1.880426	-1.629273
diff		-.312389	.1315951		-.5708356	-.0539424

diff = mean(**No**) - mean(**Yes**) t = -2.3739

Ho: diff = 0 degrees of freedom = 596

Ha: diff < 0	Ha: diff != 0	Ha: diff > 0
Pr(T < t) = 0.0090	Pr(|T| > |t|) = 0.0179	Pr(T > t) = 0.9910

- Which group has a higher average haz-score?
  - *Children aged 2 to 5 years who eat meat, poultry or fish in the last 24 hours have higher haz (-1.64 compared to -1.95)*

3. T-Tests

- Code: ttest haz, by(meat_poul_fish_C)

```
. ttest haz, by( meat_poul_fish_C)

Two-sample t test with equal variances

   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |     224   -1.950223    .0990268    1.482097   -2.145371   -1.755075
     Yes |     374   -1.637834    .0827858    1.601001    -1.80062   -1.475049
---------+--------------------------------------------------------------------
combined |     598   -1.754849    .0639408     1.56361   -1.880426   -1.629273
---------+--------------------------------------------------------------------
    diff |             -.312389    .1315951               -.5708356   -.0539424
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t =  -2.3739
Ho: diff = 0                                     degrees of freedom =      596

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0090         Pr(|T| > |t|) = 0.0179          Pr(T > t) = 0.9910
```

- Standard error, standard deviation, and 95% confidence interval of the haz of each group

3. T-Tests

- Code: ttest haz, by(meat_poul_fish_C)

```
. ttest haz, by( meat_poul_fish_C)

Two-sample t test with equal variances

   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |     224   -1.950223    .0990268    1.482097   -2.145371   -1.755075
     Yes |     374   -1.637834    .0827858    1.601001    -1.80062   -1.475049
---------+--------------------------------------------------------------------
combined |     598   -1.754849    .0639408     1.56361   -1.880426   -1.629273
---------+--------------------------------------------------------------------
    diff |            -.312389    .1315951               -.5708356   -.0539424
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t =  -2.3739
Ho: diff = 0                                     degrees of freedom =      596

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0090         Pr(|T| > |t|) = 0.0179          Pr(T > t) = 0.9910
```

- Do the two confidence intervals overlap?

3. T-Tests

- Code: ttest haz, by(meat_poul_fish_C)

```
. ttest haz, by( meat_poul_fish_C)

Two-sample t test with equal variances
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
No	224	-1.950223	.0990268	1.482097	-2.145371	-1.755075
Yes	374	-1.637834	.0827858	1.601001	-1.80062	-1.475049
combined	598	-1.754849	.0639408	1.56361	-1.880426	-1.629273
diff		-.312389	.1315951		-.5708356	-.0539424

```
    diff = mean(No) - mean(Yes)                                   t =  -2.3739
Ho: diff = 0                                     degrees of freedom =      596

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0090         Pr(|T| > |t|) = 0.0179          Pr(T > t) = 0.9910
```

- Do the two confidence intervals overlap?
  - *No - the max for "no" is -1.755 and the min for "yes" is -1.801*

3. T-Tests

- Code: ttest haz, by(meat_poul_fish_C)

```
. ttest haz, by( meat_poul_fish_C)

Two-sample t test with equal variances

------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |     224   -1.950223    .0990268    1.482097   -2.145371   -1.755075
     Yes |     374   -1.637834    .0827858    1.601001    -1.80062   -1.475049
---------+--------------------------------------------------------------------
combined |     598   -1.754849    .0639408     1.56361   -1.880426   -1.629273
---------+--------------------------------------------------------------------
    diff |           -.312389    .1315951               -.5708356   -.0539424
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t =  -2.3739
Ho: diff = 0                                     degrees of freedom =      596

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0090         Pr(|T| > |t|) = 0.0179          Pr(T > t) = 0.9910
```

- T-statistic and the degrees of freedom

3. T-Tests

- Code: ttest haz, by(meat_poul_fish_C)

```
. ttest haz, by( meat_poul_fish_C)

Two-sample t test with equal variances

   Group       Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]

      No       224   -1.950223    .0990268    1.482097   -2.145371   -1.755075
     Yes       374   -1.637834    .0827858    1.601001    -1.80062   -1.475049

combined       598   -1.754849    .0639408     1.56361   -1.880426   -1.629273

    diff             -.312389    .1315951               -.5708356   -.0539424

    diff = mean(No) - mean(Yes)                                   t =  -2.3739
Ho: diff = 0                                     degrees of freedom =      596

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0090         Pr(|T| > |t|) = 0.0179          Pr(T > t) = 0.9910
```

- 3 different p-values

3. T-Tests

- Code: ttest haz, by(meat_poul_fish_C)

```
. ttest haz, by( meat_poul_fish_C)

Two-sample t test with equal variances

   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |     224   -1.950223    .0990268    1.482097   -2.145371   -1.755075
     Yes |     374   -1.637834    .0827858    1.601001    -1.80062   -1.475049
---------+--------------------------------------------------------------------
combined |     598   -1.754849    .0639408     1.56361   -1.880426   -1.629273
---------+--------------------------------------------------------------------
    diff |           -.312389    .1315951               -.5708356   -.0539424
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t =  -2.3739
Ho: diff = 0                                     degrees of freedom =      596

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0090         Pr(|T| > |t|) = 0.0179          Pr(T > t) = 0.9910
```

- P-value for whether the mean of Group 1 is less than the mean of Group 2

3. T-Tests

- Code: ttest haz, by(meat_poul_fish_C)

```
. ttest haz, by( meat_poul_fish_C)

Two-sample t test with equal variances

   Group      Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]

      No      224   -1.950223    .0990268    1.482097   -2.145371   -1.755075
     Yes      374   -1.637834    .0827858    1.601001    -1.80062   -1.475049

combined      598   -1.754849    .0639408     1.56361   -1.880426   -1.629273

    diff            -.312389    .1315951               -.5708356   -.0539424

    diff = mean(No) - mean(Yes)                                   t =  -2.3739
Ho: diff = 0                                     degrees of freedom =      596

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0090         Pr(|T| > |t|) = 0.0179          Pr(T > t) = 0.9910
```

- P-value for whether the mean of Group 1 is not equal to the mean of Group 2

3. T-Tests

- Code: ttest haz, by(meat_poul_fish_C)

```
. ttest haz, by( meat_poul_fish_C)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |     224   -1.950223    .0990268    1.482097   -2.145371   -1.755075
     Yes |     374   -1.637834    .0827858    1.601001    -1.80062   -1.475049
---------+--------------------------------------------------------------------
combined |     598   -1.754849    .0639408     1.56361   -1.880426   -1.629273
---------+--------------------------------------------------------------------
    diff |           -.312389    .1315951               -.5708356   -.0539424
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t =  -2.3739
Ho: diff = 0                                     degrees of freedom =      596

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0090         Pr(|T| > |t|) = 0.0179          Pr(T > t) = 0.9910
```

- P-value for whether the mean of Group 1 is greater than the mean of Group 2

3. T-Tests

- Code: ttest haz, by(meat_poul_fish_C)

```
. ttest haz, by( meat_poul_fish_C)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |     224   -1.950223    .0990268    1.482097   -2.145371   -1.755075
     Yes |     374   -1.637834    .0827858    1.601001    -1.80062   -1.475049
---------+--------------------------------------------------------------------
combined |     598   -1.754849    .0639408     1.56361   -1.880426   -1.629273
---------+--------------------------------------------------------------------
    diff |            -.312389    .1315951               -.5708356   -.0539424
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t =  -2.3739
Ho: diff = 0                                     degrees of freedom =      596

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0090         Pr(|T| > |t|) = 0.0179          Pr(T > t) = 0.9910
```

- What is the p-value for whether the mean of Group 1 is less than the mean of Group 2

3. T-Tests

- Code: ttest haz, by(meat_poul_fish_C)

```
. ttest haz, by( meat_poul_fish_C)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |     224   -1.950223    .0990268    1.482097   -2.145371   -1.755075
     Yes |     374   -1.637834    .0827858    1.601001    -1.80062   -1.475049
---------+--------------------------------------------------------------------
combined |     598   -1.754849    .0639408     1.56361   -1.880426   -1.629273
---------+--------------------------------------------------------------------
    diff |            -.312389    .1315951               -.5708356   -.0539424
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t =  -2.3739
Ho: diff = 0                                     degrees of freedom =      596

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0090         Pr(|T| > |t|) = 0.0179          Pr(T > t) = 0.9910
```

- What is the p-value for whether the mean of Group 1 is less than the mean of Group 2
  - *P-value is 0.0090*

3. T-Tests

- Code: ttest haz, by(meat_poul_fish_C)

```
. ttest haz, by( meat_poul_fish_C)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |     224   -1.950223    .0990268    1.482097   -2.145371   -1.755075
     Yes |     374   -1.637834    .0827858    1.601001    -1.80062   -1.475049
---------+--------------------------------------------------------------------
combined |     598   -1.754849    .0639408     1.56361   -1.880426   -1.629273
---------+--------------------------------------------------------------------
    diff |            -.312389    .1315951               -.5708356   -.0539424
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t =  -2.3739
Ho: diff = 0                                     degrees of freedom =      596

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0090         Pr(|T| > |t|) = 0.0179          Pr(T > t) = 0.9910
```

- This means that we can say with more than 99% confidence that the mean height-for-age z-score of children aged 2 to 5 years who **eat** meat, poultry or fish in the last 24 hours is larger than the mean height-for-age z- score of children aged 2 to 5 years who **do not eat** meat, poultry or fish in the last 24 hours.

3. T-Tests

- Now let’s see if children under 5 year belonging to households with a youth household head have significantly different child height-for-age z-score than children under 5 years belonging to households with older household heads.

3. T-Tests

- Now let's see if children under 5 year belonging to households with a youth household head have significantly different child height-for-age z-score than children under 5 years belonging to households with older household heads.

- <u>Code:</u>

3. T-Tests

- Now let's see if children under 5 year belonging to households with a youth household head have significantly different child height-for-age z-score than children under 5 years belonging to households with older household heads.

- Code: *ttest haz, by(hhh_mature)*

3. T-Tests

- *ttest haz, by(hhh_mature)*

3. T-Tests

- *ttest haz, by(hhh_mature)*

```
. ttest haz, by( hhh_mature )

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
Youth he |     411   -1.435815     .085222    1.727717   -1.603342   -1.268289
Mature h |     894   -1.317058    .0589603    1.762903   -1.432775   -1.201341
---------+--------------------------------------------------------------------
combined |   1,305    -1.35446    .0485015    1.752105   -1.449609    -1.25931
---------+--------------------------------------------------------------------
    diff |           -.1187569    .1044063               -.3235798     .086066
------------------------------------------------------------------------------
    diff = mean(Youth he) - mean(Mature h)                        t =  -1.1374
Ho: diff = 0                                     degrees of freedom =     1303

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.1278         Pr(|T| > |t|) = 0.2556          Pr(T > t) = 0.8722
```

3. T-Tests

- *ttest haz, by(hhh_mature)*

```
. ttest haz, by( hhh_mature )

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
Youth he |     411   -1.435815     .085222    1.727717   -1.603342   -1.268289
Mature h |     894   -1.317058    .0589603    1.762903   -1.432775   -1.201341
---------+--------------------------------------------------------------------
combined |   1,305    -1.35446    .0485015    1.752105   -1.449609    -1.25931
---------+--------------------------------------------------------------------
    diff |           -.1187569    .1044063               -.3235798     .086066
------------------------------------------------------------------------------
    diff = mean(Youth he) - mean(Mature h)                        t =  -1.1374
Ho: diff = 0                                     degrees of freedom =     1303

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.1278         Pr(|T| > |t|) = 0.2556          Pr(T > t) = 0.8722
```

- Which mean is higher?

3. T-Tests

- *ttest haz, by(hhh_mature)*

```
. ttest haz, by( hhh_mature )

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
Youth he |     411   -1.435815     .085222    1.727717   -1.603342   -1.268289
Mature h |     894   -1.317058    .0589603    1.762903   -1.432775   -1.201341
---------+--------------------------------------------------------------------
combined |   1,305    -1.35446    .0485015    1.752105   -1.449609    -1.25931
---------+--------------------------------------------------------------------
    diff |           -.1187569    .1044063               -.3235798     .086066
------------------------------------------------------------------------------
    diff = mean(Youth he) - mean(Mature h)                        t =  -1.1374
Ho: diff = 0                                     degrees of freedom =     1303

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.1278         Pr(|T| > |t|) = 0.2556          Pr(T > t) = 0.8722
```

- Which mean is higher? *Children under 5 years belonging to a mature headed households have higher average haz-score*

3. T-Tests

- *ttest haz, by(hhh_mature)*

```
. ttest haz, by( hhh_mature )

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
Youth he |     411   -1.435815     .085222    1.727717   -1.603342   -1.268289
Mature h |     894   -1.317058    .0589603    1.762903   -1.432775   -1.201341
---------+--------------------------------------------------------------------
combined |   1,305    -1.35446    .0485015    1.752105   -1.449609    -1.25931
---------+--------------------------------------------------------------------
    diff |           -.1187569    .1044063               -.3235798     .086066
------------------------------------------------------------------------------
    diff = mean(Youth he) - mean(Mature h)                        t =  -1.1374
Ho: diff = 0                                     degrees of freedom =     1303

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.1278         Pr(|T| > |t|) = 0.2556          Pr(T > t) = 0.8722
```

- What is the p-value that the mean of children under 5 years coming from youth headed households is less than that of children under 5 years coming from mature headed households? Is it significant?

3. T-Tests

- *ttest haz, by(hhh_mature)*

```
. ttest haz, by( hhh_mature )

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
Youth he |     411   -1.435815     .085222    1.727717   -1.603342   -1.268289
Mature h |     894   -1.317058    .0589603    1.762903   -1.432775   -1.201341
---------+--------------------------------------------------------------------
combined |   1,305    -1.35446    .0485015    1.752105   -1.449609    -1.25931
---------+--------------------------------------------------------------------
    diff |           -.1187569    .1044063               -.3235798     .086066
------------------------------------------------------------------------------
    diff = mean(Youth he) - mean(Mature h)                        t =  -1.1374
Ho: diff = 0                                     degrees of freedom =     1303

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.1278         Pr(|T| > |t|) = 0.2556          Pr(T > t) = 0.8722
```

- What is the p-value that the mean of children under 5 years coming from youth headed households is less than that of children under 5 years coming from mature headed households? Is it significant?
  - *P-value=0.1278 – it is not significant at the 90% level or above*

4. Ordinary least-squares (OLS) linear regression

- Now we are going to move on to our final analysis – an ordinary least-squares (OLS) linear regression

4. Ordinary least-squares (OLS) linear regression

- Now we are going to move on to our final analysis – an ordinary least-squares (OLS) linear regression

- Let’s just start with one independent variable (the one that we used in our correlations – ex_r)

4. Ordinary least-squares (OLS) linear regression

- Now we are going to move on to our final analysis – an ordinary least-squares (OLS) linear regression

- Let's just start with one independent variable (the one that we used in our correlations – ex_r)

- Code: *regress haz ex_r*

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r*

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r*

```
. regress haz ex_r

      Source |       SS           df       MS      Number of obs   =     1,305
-------------+----------------------------------   F(1, 1303)      =     18.84
       Model |  57.052925         1  57.052925   Prob > F        =    0.0000
    Residual |  3946.05972     1,303  3.02844184   R-squared       =    0.0143
-------------+----------------------------------   Adj R-squared   =    0.0135
       Total |  4003.11264     1,304  3.06987166   Root MSE        =    1.7402

------------------------------------------------------------------------------
         haz |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ex_r |   .0052426   .0012079     4.34   0.000      .002873    .0076121
       _cons |  -1.650508    .083504   -19.77   0.000    -1.814325   -1.486691
```

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r*

```
. regress haz ex_r

      Source |       SS           df       MS      Number of obs   =     1,305
-------------+----------------------------------   F(1, 1303)      =     18.84
       Model |   57.052925         1   57.052925   Prob > F        =    0.0000
    Residual |  3946.05972     1,303  3.02844184   R-squared       =    0.0143
-------------+----------------------------------   Adj R-squared   =    0.0135
       Total |  4003.11264     1,304  3.06987166   Root MSE        =    1.7402

------------------------------------------------------------------------------
         haz |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ex_r |   .0052426   .0012079     4.34   0.000      .002873    .0076121
       _cons |  -1.650508    .083504   -19.77   0.000    -1.814325   -1.486691
```

- So what does this output show us?

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r*

```
. regress haz ex_r

      Source |       SS           df       MS      Number of obs   =     1,305
-------------+----------------------------------   F(1, 1303)      =     18.84
       Model |  57.052925         1   57.052925   Prob > F        =    0.0000
    Residual |  3946.05972     1,303  3.02844184   R-squared       =    0.0143
-------------+----------------------------------   Adj R-squared   =    0.0135
       Total |  4003.11264     1,304  3.06987166   Root MSE        =    1.7402

------------------------------------------------------------------------------
         haz |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ex_r |   .0052426   .0012079     4.34   0.000      .002873    .0076121
       _cons |  -1.650508    .083504   -19.77   0.000    -1.814325   -1.486691
```

- *Anova Table* – sum of squares, degrees of freedom, and mean squares

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r*

```
. regress haz ex_r

      Source |       SS           df       MS      Number of obs   =     1,305
-------------+----------------------------------   F(1, 1303)      =     18.84
       Model |   57.052925         1   57.052925   Prob > F        =    0.0000
    Residual |  3946.05972     1,303  3.02844184   R-squared       =    0.0143
-------------+----------------------------------   Adj R-squared   =    0.0135
       Total |  4003.11264     1,304  3.06987166   Root MSE        =    1.7402

------------------------------------------------------------------------------
         haz |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ex_r |   .0052426   .0012079     4.34   0.000      .002873    .0076121
       _cons |  -1.650508    .083504   -19.77   0.000    -1.814325   -1.486691
```

- *Overall model fit* - number of observations, F-statistic, R-squared, etc.

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r*

```
. regress haz ex_r

      Source |       SS           df       MS      Number of obs   =     1,305
-------------+----------------------------------   F(1, 1303)      =     18.84
       Model |   57.052925         1   57.052925   Prob > F        =    0.0000
    Residual |  3946.05972     1,303  3.02844184   R-squared       =    0.0143
-------------+----------------------------------   Adj R-squared   =    0.0135
       Total |  4003.11264     1,304  3.06987166   Root MSE        =    1.7402

------------------------------------------------------------------------------
         haz |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ex_r |   .0052426   .0012079     4.34   0.000      .002873    .0076121
       _cons |  -1.650508    .083504   -19.77   0.000    -1.814325   -1.486691
------------------------------------------------------------------------------
```

- *Parameter estimates* – coefficient, standard error, t-statistic, p-value, and 95% confidence interval for all independent variables (right now we only have 1 IV)

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r*

```
. regress haz ex_r

      Source |       SS           df       MS      Number of obs   =     1,305
-------------+----------------------------------   F(1, 1303)      =     18.84
       Model |   57.052925         1   57.052925   Prob > F        =    0.0000
    Residual |  3946.05972     1,303  3.02844184   R-squared       =    0.0143
-------------+----------------------------------   Adj R-squared   =    0.0135
       Total |  4003.11264     1,304  3.06987166   Root MSE        =    1.7402

------------------------------------------------------------------------------
         haz |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ex_r |   .0052426   .0012079     4.34   0.000      .002873    .0076121
       _cons |  -1.650508    .083504   -19.77   0.000    -1.814325   -1.486691
```

- Does this p-value look familiar at all?

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r*

```
. regress haz ex_r

      Source |       SS           df       MS      Number of obs   =     1,305
-------------+----------------------------------   F(1, 1303)      =     18.84
       Model |   57.052925         1   57.052925   Prob > F        =    0.0000
    Residual |  3946.05972     1,303  3.02844184   R-squared       =    0.0143
-------------+----------------------------------   Adj R-squared   =    0.0135
       Total |  4003.11264     1,304  3.06987166   Root MSE        =    1.7402

------------------------------------------------------------------------------
         haz |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ex_r |   .0052426   .0012079     4.34   0.000      .002873    .0076121
       _cons |  -1.650508    .083504   -19.77   0.000    -1.814325   -1.486691
```

- Does this p-value look familiar at all? *It's the same as from our correlation (pwcorr haz ex_r, sig)*

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r*

```
. regress haz ex_r

      Source |       SS           df       MS      Number of obs   =     1,305
-------------+----------------------------------   F(1, 1303)      =     18.84
       Model |   57.052925         1   57.052925   Prob > F        =    0.0000
    Residual |  3946.05972     1,303  3.02844184   R-squared       =    0.0143
-------------+----------------------------------   Adj R-squared   =    0.0135
       Total |  4003.11264     1,304  3.06987166   Root MSE        =    1.7402

------------------------------------------------------------------------------
         haz |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ex_r |   .0052426   .0012079     4.34   0.000      .002873    .0076121
       _cons |  -1.650508    .083504   -19.77   0.000    -1.814325   -1.486691
```

```
. pwcorr haz ex_r, sig

             |      haz     ex_r
-------------+------------------
         haz |   1.0000
             |
             |
        ex_r |   0.1194   1.0000
             |   0.0000
```

- Does this p-value look familiar at all? *It's the same as from our correlation (pwcorr haz ex_r, sig)*
    - *A regression with only 1 independent variable is essentially a correlation*

4. Ordinary least-squares (OLS) linear regression

- Now, let's try adding some more variables

4. Ordinary least-squares (OLS) linear regression

- Now, let's try adding some more variables

- What else might be associated with child height-for-age z-score?

4. Ordinary least-squares (OLS) linear regression

- Now, let's try adding some more variables

- What else might be associated with household dietary diversity?

- Let's use ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature and hhh_female as our independent variables (IVs).

4. Ordinary least-squares (OLS) linear regression

- Now, let’s try adding some more variables
- What else might be associated with household dietary diversity?
- Let’s use ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature and hhh_female as our independent variables (IVs).
- Code:

4. Ordinary least-squares (OLS) linear regression

- Now, let's try adding some more variables
- What else might be associated with household dietary diversity?
- Let's use ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature and hhh_female as our independent variables (IVs).
- Code: *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female*

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female*

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female*
- What types of variables are the IVs in this regression?

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female*
- What types of variables are the IVs in this regression?

codebook ex_r
codebook safewater
codebook improvedtoilets_vs1
codebook advice2
codebook landholdings
codebook hhh_mature
codebook hhh_female

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female*
- What types of variables are the IVs in this regression?

  codebook ex_r – numeric/continuous

  codebook safewater – categorical dummy variable (0/1)

  codebook improvedtoilets_vs1 - categorical dummy variable (0/1)

  codebook advice2 - *categorical dummy variable (0/1)*

  codebook landholdings – numeric/continuous

  codebook hhh_mature - categorical dummy variable (0/1)

  codebook hhh_female - categorical dummy variable (0/1)

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female*
- What types of variables are the IVs in this regression?
  - *codebook ex_r* – numeric/continuous
  - *codebook safewater* – categorical dummy variable (0/1)
  - *codebook improvedtoilets_vs1* - categorical dummy variable (0/1)
  - *codebook advice2* - *categorical dummy variable (0/1)*
  - *codebook landholdings* – numeric/continuous
  - *codebook hhh_mature* - categorical dummy variable (0/1)
  - *codebook hhh_female* - categorical dummy variable (0/1)
- All of these variable types are ready to go straight into the regression

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female*

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female*

```
. regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female

      Source |       SS           df       MS      Number of obs   =     1,305
-------------+----------------------------------   F(7, 1297)      =      5.75
       Model |  120.572361         7  17.2246229   Prob > F        =    0.0000
    Residual |  3882.54028     1,297  2.99347747   R-squared       =    0.0301
-------------+----------------------------------   Adj R-squared   =    0.0249
       Total |  4003.11264     1,304  3.06987166   Root MSE        =    1.7302

------------------------------------------------------------------------------
         haz | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        ex_r |   .0046134   .0012492     3.69   0.000     .0021627    .0070641
   safewater |   .0789408   .1091627     0.72   0.470     -.135214    .2930956
improvedtoilets_vs1 |   .3604228    .100654     3.58   0.000     .1629602    .5578853
     advice2 |   .1907452   .0968147     1.97   0.049     .0008146    .3806757
landholdings |   -.065619   .0351863    -1.86   0.062    -.1346472    .0034093
  hhh_mature |   .0412775   .1060595     0.39   0.697    -.1667895    .2493445
  hhh_female |    .010044   .1911317     0.05   0.958    -.3649172    .3850051
       _cons |  -1.800801   .1246846   -14.44   0.000    -2.045406   -1.556195
------------------------------------------------------------------------------
```

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female*

```
. regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female
```

Source	SS	df	MS
Model	120.572361	7	17.2246229
Residual	3882.54028	1,297	2.99347747
Total	4003.11264	1,304	3.06987166

Number of obs	=	1,305
F(7, 1297)	=	5.75
Prob > F	=	0.0000
R-squared	=	0.0301
Adj R-squared	=	0.0249
Root MSE	=	1.7302

haz	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
ex_r	.0046134	.0012492	3.69	0.000	.0021627	.0070641
safewater	.0789408	.1091627	0.72	0.470	-.135214	.2930956
improvedtoilets_vs1	.3604228	.100654	3.58	0.000	.1629602	.5578853
advice2	.1907452	.0968147	1.97	0.049	.0008146	.3806757
landholdings	-.065619	.0351863	-1.86	0.062	-.1346472	.0034093
hhh_mature	.0412775	.1060595	0.39	0.697	-.1667895	.2493445
hhh_female	.010044	.1911317	0.05	0.958	-.3649172	.3850051
_cons	-1.800801	.1246846	-14.44	0.000	-2.045406	-1.556195

- Are any of these variables significant?

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female*

```
. regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female

      Source |       SS           df       MS      Number of obs   =     1,305
-------------+----------------------------------   F(7, 1297)      =      5.75
       Model |  120.572361         7  17.2246229   Prob > F        =    0.0000
    Residual |  3882.54028     1,297  2.99347747   R-squared       =    0.0301
-------------+----------------------------------   Adj R-squared   =    0.0249
       Total |  4003.11264     1,304  3.06987166   Root MSE        =    1.7302

------------------------------------------------------------------------------------
               haz | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------------+----------------------------------------------------------------
              ex_r |   .0046134   .0012492     3.69   0.000     .0021627    .0070641
         safewater |   .0789408   .1091627     0.72   0.470     -.135214    .2930956
improvedtoilets_vs1|   .3604228    .100654     3.58   0.000     .1629602    .5578853
           advice2 |   .1907452   .0968147     1.97   0.049     .0008146    .3806757
      landholdings |   -.065619   .0351863    -1.86   0.062    -.1346472    .0034093
        hhh_mature |   .0412775   .1060595     0.39   0.697    -.1667895    .2493445
        hhh_female |    .010044   .1911317     0.05   0.958    -.3649172    .3850051
             _cons |  -1.800801   .1246846   -14.44   0.000    -2.045406   -1.556195
------------------------------------------------------------------------------------
```

- Are any of these variables significant?
  - *ex_r is significant at the .01 level*
  - *safewater is significant at the .05 level*
  - *improvedtoilets_vs1 is significant at the .01 level*
  - *ddvice2 is significant at the .05 level*

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female*

```
. regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female

      Source |       SS           df       MS      Number of obs   =     1,305
-------------+----------------------------------   F(7, 1297)      =      5.75
       Model |  120.572361         7  17.2246229   Prob > F        =    0.0000
    Residual |  3882.54028     1,297  2.99347747   R-squared       =    0.0301
-------------+----------------------------------   Adj R-squared   =    0.0249
       Total |  4003.11264     1,304  3.06987166   Root MSE        =    1.7302

------------------------------------------------------------------------------
         haz | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        ex_r |   .0046134   .0012492     3.69   0.000     .0021627    .0070641
   safewater |   .0789408   .1091627     0.72   0.470     -.135214    .2930956
improvedtoilets_vs1 |   .3604228    .100654     3.58   0.000     .1629602    .5578853
     advice2 |   .1907452   .0968147     1.97   0.049     .0008146    .3806757
landholdings |   -.065619   .0351863    -1.86   0.062    -.1346472    .0034093
  hhh_mature |   .0412775   .1060595     0.39   0.697    -.1667895    .2493445
  hhh_female |    .010044   .1911317     0.05   0.958    -.3649172    .3850051
       _cons |  -1.800801   .1246846   -14.44   0.000    -2.045406   -1.556195
------------------------------------------------------------------------------
```

- How can we interpret the coefficient on ex_r, knowing that it's significant?

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female*

```
. regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female

      Source |       SS           df       MS      Number of obs   =     1,305
-------------+----------------------------------   F(7, 1297)      =      5.75
       Model |  120.572361         7  17.2246229   Prob > F        =    0.0000
    Residual |  3882.54028     1,297  2.99347747   R-squared       =    0.0301
-------------+----------------------------------   Adj R-squared   =    0.0249
       Total |  4003.11264     1,304  3.06987166   Root MSE        =    1.7302

------------------------------------------------------------------------------
         haz | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        ex_r |   .0046134   .0012492     3.69   0.000     .0021627    .0070641
   safewater |   .0789408   .1091627     0.72   0.470     -.135214    .2930956
improvedtoilets_vs1 |   .3604228    .100654     3.58   0.000     .1629602    .5578853
     advice2 |   .1907452   .0968147     1.97   0.049     .0008146    .3806757
landholdings |   -.065619   .0351863    -1.86   0.062    -.1346472    .0034093
  hhh_mature |   .0412775   .1060595     0.39   0.697    -.1667895    .2493445
  hhh_female |    .010044   .1911317     0.05   0.958    -.3649172    .3850051
       _cons |  -1.800801   .1246846   -14.44   0.000    -2.045406   -1.556195
------------------------------------------------------------------------------
```

- *Because ex-r is a continuous variable, the coefficient says that the marginal effect of one additional increase in daily real consumption expenditure increases the child's haz-score by 0.0046. (The coefficient is 0.005), keeping all other factors constant.*

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female*

```
. regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female
```

Source	SS	df	MS
Model	120.572361	7	17.2246229
Residual	3882.54028	1,297	2.99347747
Total	4003.11264	1,304	3.06987166

Number of obs	=	1,305
F(7, 1297)	=	5.75
Prob > F	=	0.0000
R-squared	=	0.0301
Adj R-squared	=	0.0249
Root MSE	=	1.7302

haz	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
ex_r	.0046134	.0012492	3.69	0.000	.0021627	.0070641
safewater	.0789408	.1091627	0.72	0.470	-.135214	.2930956
improvedtoilets_vs1	.3604228	.100654	3.58	0.000	.1629602	.5578853
advice2	.1907452	.0968147	1.97	0.049	.0008146	.3806757
landholdings	-.065619	.0351863	-1.86	0.062	-.1346472	.0034093
hhh_mature	.0412775	.1060595	0.39	0.697	-.1667895	.2493445
hhh_female	.010044	.1911317	0.05	0.958	-.3649172	.3850051
_cons	-1.800801	.1246846	-14.44	0.000	-2.045406	-1.556195

- How can we interpret the coefficient on improvedtoilets_vs1, knowing that it's significant?

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female*

```
. regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female

      Source |       SS           df       MS      Number of obs   =     1,305
-------------+----------------------------------   F(7, 1297)      =      5.75
       Model |  120.572361         7  17.2246229   Prob > F        =    0.0000
    Residual |  3882.54028     1,297  2.99347747   R-squared       =    0.0301
-------------+----------------------------------   Adj R-squared   =    0.0249
       Total |  4003.11264     1,304  3.06987166   Root MSE        =    1.7302

------------------------------------------------------------------------------
         haz | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        ex_r |   .0046134   .0012492     3.69   0.000     .0021627    .0070641
   safewater |   .0789408   .1091627     0.72   0.470     -.135214    .2930956
improvedtoilets_vs1 |   .3604228    .100654     3.58   0.000     .1629602    .5578853
     advice2 |   .1907452   .0968147     1.97   0.049     .0008146    .3806757
landholdings |   -.065619   .0351863    -1.86   0.062    -.1346472    .0034093
  hhh_mature |   .0412775   .1060595     0.39   0.697    -.1667895    .2493445
  hhh_female |    .010044   .1911317     0.05   0.958    -.3649172    .3850051
       _cons |  -1.800801   .1246846   -14.44   0.000    -2.045406   -1.556195
------------------------------------------------------------------------------
```

- *Because improvetoilets_vs1 is a dummy variable, the coefficient says that if children under 5 years belong to a households that have access to improved toilet infrastructures, the haz-score will <u>increase</u> by 0.36. (The coefficient is 0.36).*

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female*

```
. regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female

      Source |       SS           df       MS      Number of obs   =     1,305
-------------+----------------------------------   F(7, 1297)      =      5.75
       Model |  120.572361         7  17.2246229   Prob > F        =    0.0000
    Residual |  3882.54028     1,297  2.99347747   R-squared       =    0.0301
-------------+----------------------------------   Adj R-squared   =    0.0249
       Total |  4003.11264     1,304  3.06987166   Root MSE        =    1.7302

------------------------------------------------------------------------------
         haz | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        ex_r |   .0046134   .0012492     3.69   0.000     .0021627    .0070641
   safewater |   .0789408   .1091627     0.72   0.470     -.135214    .2930956
improvedtoilets_vs1 |   .3604228    .100654     3.58   0.000     .1629602    .5578853
     advice2 |   .1907452   .0968147     1.97   0.049     .0008146    .3806757
landholdings |   -.065619   .0351863    -1.86   0.062    -.1346472    .0034093
  hhh_mature |   .0412775   .1060595     0.39   0.697    -.1667895    .2493445
  hhh_female |    .010044   .1911317     0.05   0.958    -.3649172    .3850051
       _cons |  -1.800801   .1246846   -14.44   0.000    -2.045406   -1.556195
------------------------------------------------------------------------------
```

- How can we interpret the coefficient on safewater, knowing that it’s significant? (with 95% confidence)

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female*

```
. regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female

      Source |       SS           df       MS      Number of obs   =     1,305
-------------+----------------------------------   F(7, 1297)      =      5.75
       Model |  120.572361         7  17.2246229   Prob > F        =    0.0000
    Residual |  3882.54028     1,297  2.99347747   R-squared       =    0.0301
-------------+----------------------------------   Adj R-squared   =    0.0249
       Total |  4003.11264     1,304  3.06987166   Root MSE        =    1.7302

------------------------------------------------------------------------------
         haz | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        ex_r |   .0046134   .0012492     3.69   0.000     .0021627    .0070641
   safewater |   .0789408   .1091627     0.72   0.470     -.135214    .2930956
improvedtoilets_vs1 | .3604228 .100654   3.58   0.000     .1629602    .5578853
     advice2 |   .1907452   .0968147     1.97   0.049     .0008146    .3806757
landholdings |   -.065619   .0351863    -1.86   0.062    -.1346472    .0034093
  hhh_mature |   .0412775   .1060595     0.39   0.697    -.1667895    .2493445
  hhh_female |    .010044   .1911317     0.05   0.958    -.3649172    .3850051
       _cons |  -1.800801   .1246846   -14.44   0.000    -2.045406   -1.556195
------------------------------------------------------------------------------
```

- *Because safewater is a dummy variable, the coefficient says that if children under 5 years belong to a household that treats their water before drinking, their haz-score will <u>increase</u> by 0.079.*

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female*

```
. regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female

      Source |       SS           df       MS      Number of obs   =     1,305
-------------+----------------------------------   F(7, 1297)      =      5.75
       Model |  120.572361         7  17.2246229   Prob > F        =    0.0000
    Residual |  3882.54028     1,297  2.99347747   R-squared       =    0.0301
-------------+----------------------------------   Adj R-squared   =    0.0249
       Total |  4003.11264     1,304  3.06987166   Root MSE        =    1.7302

------------------------------------------------------------------------------
         haz | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        ex_r |   .0046134   .0012492     3.69   0.000     .0021627    .0070641
   safewater |   .0789408   .1091627     0.72   0.470     -.135214    .2930956
improvedtoilets_vs1 |   .3604228    .100654     3.58   0.000     .1629602    .5578853
     advice2 |   .1907452   .0968147     1.97   0.049     .0008146    .3806757
landholdings |   -.065619   .0351863    -1.86   0.062    -.1346472    .0034093
  hhh_mature |   .0412775   .1060595     0.39   0.697    -.1667895    .2493445
  hhh_female |    .010044   .1911317     0.05   0.958    -.3649172    .3850051
       _cons |  -1.800801   .1246846   -14.44   0.000    -2.045406   -1.556195
------------------------------------------------------------------------------
```

- How can we interpret the coefficient on advice2, knowing that it's significant? (with 95% confidence)

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female*

```
. regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female

      Source |       SS           df       MS      Number of obs   =     1,305
-------------+----------------------------------   F(7, 1297)      =      5.75
       Model |  120.572361         7  17.2246229   Prob > F        =    0.0000
    Residual |  3882.54028     1,297  2.99347747   R-squared       =    0.0301
-------------+----------------------------------   Adj R-squared   =    0.0249
       Total |  4003.11264     1,304  3.06987166   Root MSE        =    1.7302

------------------------------------------------------------------------------
         haz | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        ex_r |   .0046134   .0012492     3.69   0.000     .0021627    .0070641
   safewater |   .0789408   .1091627     0.72   0.470     -.135214    .2930956
improvedtoilets_vs1 |   .3604228    .100654     3.58   0.000     .1629602    .5578853
     advice2 |   .1907452   .0968147     1.97   0.049     .0008146    .3806757
landholdings |   -.065619   .0351863    -1.86   0.062    -.1346472    .0034093
  hhh_mature |   .0412775   .1060595     0.39   0.697    -.1667895    .2493445
  hhh_female |    .010044   .1911317     0.05   0.958    -.3649172    .3850051
       _cons |  -1.800801   .1246846   -14.44   0.000    -2.045406   -1.556195
------------------------------------------------------------------------------
```

- *Because advice2 is a dummy variable, the coefficient says that if children under 5 years belonged to a household that received extension on appropriate nutrition for pregnant women, their haz-score will <u>increase</u> by 0.190. (The coefficient is 0.1907452)*

4. Ordinary least-squares (OLS) linear regression

- Now, let’s add province. What type of variable is province?

4. Ordinary least-squares (OLS) linear regression

- Now, let’s add province. What type of variable is province?

- *codebook prov – categorical/numeric/byte*

4. Ordinary least-squares (OLS) linear regression

- Now, let's add province. What type of variable is province?
- *codebook prov – categorical/numeric/byte*
- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female province*

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female province*

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female province*

```
. regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female provin
> ce

      Source |       SS           df       MS      Number of obs   =     1,305
-------------+----------------------------------   F(8, 1296)      =      9.47
       Model |  221.125425         8  27.6406782   Prob > F        =    0.0000
    Residual |  3781.98722     1,296  2.91820001   R-squared       =    0.0552
-------------+----------------------------------   Adj R-squared   =    0.0494
       Total |  4003.11264     1,304  3.06987166   Root MSE        =    1.7083

------------------------------------------------------------------------------------
                haz | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
--------------------+---------------------------------------------------------------
               ex_r |   .0041256   .0012362     3.34   0.001     .0017005    .0065508
          safewater |  -.0140756     .10894    -0.13   0.897    -.2277936    .1996425
improvedtoilets_vs1 |    .327707   .0995365     3.29   0.001     .1324366    .5229774
            advice2 |   .1516445   .0958214     1.58   0.114    -.0363376    .3396266
       landholdings |  -.0137426   .0358475    -0.38   0.702     -.084068    .0565829
         hhh_mature |   .0103881   .1048496     0.10   0.921    -.1953055    .2160817
         hhh_female |    .116171   .1895773     0.61   0.540     -.255741    .4880829
           province |    .063776   .0108647     5.87   0.000     .0424617    .0850903
              _cons |  -2.290256   .1486872   -15.40   0.000     -2.58195   -1.998562
------------------------------------------------------------------------------------
```

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female province*

```
. regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female provin
> ce

      Source |       SS           df       MS      Number of obs   =     1,305
-------------+----------------------------------   F(8, 1296)      =      9.47
       Model |  221.125425         8  27.6406782   Prob > F        =    0.0000
    Residual |  3781.98722     1,296  2.91820001   R-squared       =    0.0552
-------------+----------------------------------   Adj R-squared   =    0.0494
       Total |  4003.11264     1,304  3.06987166   Root MSE        =    1.7083

------------------------------------------------------------------------------
         haz | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        ex_r |   .0041256   .0012362     3.34   0.001     .0017005    .0065508
   safewater |  -.0140756     .10894    -0.13   0.897    -.2277936    .1996425
improvedtoilets_vs1 | .327707   .0995365     3.29   0.001     .1324366    .5229774
     advice2 |   .1516445   .0958214     1.58   0.114    -.0363376    .3396266
landholdings |  -.0137426   .0358475    -0.38   0.702     -.084068    .0565829
  hhh_mature |   .0103881   .1048496     0.10   0.921    -.1953055    .2160817
  hhh_female |    .116171   .1895773     0.61   0.540     -.255741    .4880829
    province |    .063776   .0108647     5.87   0.000     .0424617    .0850903
       _cons |  -2.290256   .1486872   -15.40   0.000     -2.58195   -1.998562
------------------------------------------------------------------------------
```

- How would we interpret the coefficient on “province”?

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female province*

```
. regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female provin
> ce

      Source |       SS           df       MS      Number of obs   =     1,305
-------------+----------------------------------   F(8, 1296)      =      9.47
       Model |  221.125425         8  27.6406782   Prob > F        =    0.0000
    Residual |  3781.98722     1,296  2.91820001   R-squared       =    0.0552
-------------+----------------------------------   Adj R-squared   =    0.0494
       Total |  4003.11264     1,304  3.06987166   Root MSE        =    1.7083

------------------------------------------------------------------------------
         haz | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        ex_r |   .0041256   .0012362     3.34   0.001     .0017005    .0065508
   safewater |  -.0140756     .10894    -0.13   0.897    -.2277936    .1996425
improvedtoilets_vs1 |    .327707   .0995365     3.29   0.001     .1324366    .5229774
     advice2 |   .1516445   .0958214     1.58   0.114    -.0363376    .3396266
landholdings |  -.0137426   .0358475    -0.38   0.702     -.084068    .0565829
  hhh_mature |   .0103881   .1048496     0.10   0.921    -.1953055    .2160817
  hhh_female |    .116171   .1895773     0.61   0.540     -.255741    .4880829
    province |    .063776   .0108647     5.87   0.000     .0424617    .0850903
       _cons |  -2.290256   .1486872   -15.40   0.000     -2.58195   -1.998562
------------------------------------------------------------------------------
```

- How would we interpret the coefficient on “province”?

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female province*

```
. regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female provin
> ce

      Source |       SS           df       MS      Number of obs   =     1,305
-------------+----------------------------------   F(8, 1296)      =      9.47
       Model |  221.125425         8  27.6406782   Prob > F        =    0.0000
    Residual |  3781.98722     1,296  2.91820001   R-squared       =    0.0552
-------------+----------------------------------   Adj R-squared   =    0.0494
       Total |  4003.11264     1,304  3.06987166   Root MSE        =    1.7083

------------------------------------------------------------------------------
         haz | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        ex_r |   .0041256   .0012362     3.34   0.001     .0017005    .0065508
   safewater |  -.0140756     .10894    -0.13   0.897    -.2277936    .1996425
improvedtoilets_vs1 | .327707 .0995365  3.29   0.001     .1324366    .5229774
     advice2 |   .1516445   .0958214     1.58   0.114    -.0363376    .3396266
landholdings |  -.0137426   .0358475    -0.38   0.702     -.084068    .0565829
  hhh_mature |   .0103881   .1048496     0.10   0.921    -.1953055    .2160817
  hhh_female |    .116171   .1895773     0.61   0.540     -.255741    .4880829
    province |    .063776   .0108647     5.87   0.000     .0424617    .0850903
       _cons |  -2.290256   .1486872   -15.40   0.000     -2.58195   -1.998562
------------------------------------------------------------------------------
```

- *For every one unit increase in province, the haz-score of children under 5 years increase by 0.063…*

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female province*

```
. regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female provin
> ce
```

Source	SS	df	MS
Model	221.125425	8	27.6406782
Residual	3781.98722	1,296	2.91820001
Total	4003.11264	1,304	3.06987166

Number of obs	=	1,305
F(8, 1296)	=	9.47
Prob > F	=	0.0000
R-squared	=	0.0552
Adj R-squared	=	0.0494
Root MSE	=	1.7083

haz	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
ex_r	.0041256	.0012362	3.34	0.001	.0017005	.0065508
safewater	-.0140756	.10894	-0.13	0.897	-.2277936	.1996425
improvedtoilets_vs1	.327707	.0995365	3.29	0.001	.1324366	.5229774
advice2	.1516445	.0958214	1.58	0.114	-.0363376	.3396266
landholdings	-.0137426	.0358475	-0.38	0.702	-.084068	.0565829
hhh_mature	.0103881	.1048496	0.10	0.921	-.1953055	.2160817
hhh_female	.116171	.1895773	0.61	0.540	-.255741	.4880829
province	$\boxed{.063776}$	.0108647	5.87	0.000	.0424617	.0850903
_cons	-2.290256	.1486872	-15.40	0.000	-2.58195	-1.998562

- *For every one unit increase in province, the haz-score of children under 5 years increase by 0.063…*
- ***This doesn't make sense because province isn't a continuous variable or a dummy!***

4. Ordinary least-squares (OLS) linear regression

- To fix this, we can put “i.” in front of “province” (or any categorical variable)

4. Ordinary least-squares (OLS) linear regression

- To fix this, we can put “i.” in front of “province” (or any categorical variable)

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female i.province*

4. Ordinary least-squares (OLS) linear regression

- To fix this, we can put "i." in front of "prov" (or any categorical variable)

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female i.province*
- This now turns province into 14 dummy variables for the regression (dropping one category to avoid collinearity)

4. Ordinary least-squares (OLS) linear regression

- To fix this, we can put “i.” in front of “prov” (or any categorical variable)

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female i.province*
- This now turns province into 14 dummy variables for the regression (dropping one category to avoid collinearity)
  - Stata automatically chooses the category with the lowest value to drop (it will become the comparison group)
  - The lowest value in “province” is 1=ARoB

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female i.province*

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female i.province*

```
. regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female i.province
```

Source	SS	df	MS
Model	396.27377	20	19.8136885
Residual	3606.83887	1,284	2.80906454
Total	4003.11264	1,304	3.06987166

Number of obs	=	1,305
F(20, 1284)	=	7.05
Prob > F	=	0.0000
R-squared	=	0.0990
Adj R-squared	=	0.0850
Root MSE	=	1.676

haz	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
ex_r	.0020303	.0013157	1.54	0.123	-.0005508	.0046114
safewater	.0192869	.1093072	0.18	0.860	-.1951534	.2337272
improvedtoilets_vs1	.2902912	.1019296	2.85	0.004	.0903243	.4902581
advice2	.0962266	.1068725	0.90	0.368	-.1134372	.3058905
landholdings	.0405974	.0379195	1.07	0.285	-.0337936	.1149885
hhh_mature	-.0140334	.1038584	-0.14	0.893	-.2177842	.1897175
hhh_female	.0317556	.1871575	0.17	0.865	-.3354125	.3989236
province						
Central Province	.5943186	.2475298	2.40	0.016	.1087114	1.079926
Chimbu (Simbu) Province	.4455552	.2231952	2.00	0.046	.007688	.8834225
East New Britain Province	.9316815	.24666	3.78	0.000	.4477806	1.415582
East Sepik Province	.3376934	.2644979	1.28	0.202	-.1812021	.8565889
Eastern Highlands Province	.7891199	.2534324	3.11	0.002	.2919328	1.286307
Jiwaka Province	.5610519	.2476679	2.27	0.024	.0751739	1.04693
Madang Province	.7052731	.2313612	3.05	0.002	.2513857	1.15916
Morobe Province	-.4243679	.2527177	-1.68	0.093	-.9201529	.0714171
Western Province	1.267225	.1785732	7.10	0.000	.9168979	1.617553
Western Highlands	.5562394	.2187539	2.54	0.011	.1270851	.9853938
Gulf Province	1.293717	.2343655	5.52	0.000	.833936	1.753499
Oro Province	1.167069	.2215291	5.27	0.000	.7324708	1.601668
Milne Bay Province	1.148766	.2669933	4.30	0.000	.6249753	1.672557

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female i.province*

```
. regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female i.province

      Source |       SS           df       MS      Number of obs   =     1,305
-------------+----------------------------------   F(20, 1284)     =      7.05
       Model |   396.27377        20  19.8136885   Prob > F        =    0.0000
    Residual |  3606.83887     1,284  2.80906454   R-squared       =    0.0990
-------------+----------------------------------   Adj R-squared   =    0.0850
       Total |  4003.11264     1,304  3.06987166   Root MSE        =     1.676

------------------------------------------------------------------------------------------
                      haz | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
--------------------------+---------------------------------------------------------------
                     ex_r |   .0020303   .0013157     1.54   0.123    -.0005508    .0046114
                safewater |   .0192869   .1093072     0.18   0.860    -.1951534    .2337272
      improvedtoilets_vs1 |   .2902912   .1019296     2.85   0.004     .0903243    .4902581
                  advice2 |   .0962266   .1068725     0.90   0.368    -.1134372    .3058905
             landholdings |   .0405974   .0379195     1.07   0.285    -.0337936    .1149885
               hhh_mature |  -.0140334   .1038584    -0.14   0.893    -.2177842    .1897175
               hhh_female |   .0317556   .1871575     0.17   0.865    -.3354125    .3989236
                          |
                 province |
        Central Province  |   .5943186   .2475298     2.40   0.016     .1087114    1.079926
  Chimbu (Simbu) Province |   .4455552   .2231952     2.00   0.046      .007688    .8834225
  East New Britain Province |   .9316815     .24666     3.78   0.000     .4477806    1.415582
     East Sepik Province  |   .3376934   .2644979     1.28   0.202    -.1812021    .8565889
Eastern Highlands Province|   .7891199   .2534324     3.11   0.002     .2919328    1.286307
         Jiwaka Province  |   .5610519   .2476679     2.27   0.024     .0751739     1.04693
         Madang Province  |   .7052731   .2313612     3.05   0.002     .2513857     1.15916
         Morobe Province  |  -.4243679   .2527177    -1.68   0.093    -.9201529    .0714171
        Western Province  |   1.267225   .1785732     7.10   0.000     .9168979    1.617553
       Western Highlands  |   .5562394   .2187539     2.54   0.011     .1270851    .9853938
           Gulf Province  |   1.293717   .2343655     5.52   0.000      .833936    1.753499
            Oro Province  |   1.167069   .2215291     5.27   0.000     .7324708    1.601668
      Milne Bay Province  |   1.148766   .2669933     4.30   0.000     .6249753    1.672557
```

- Are any of the new province dummy variables significant?

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female i.province*

```
. regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female i.province
```

Source	SS	df	MS
Model	396.27377	20	19.8136885
Residual	3606.83887	1,284	2.80906454
Total	4003.11264	1,304	3.06987166

Number of obs	=	1,305
F(20, 1284)	=	7.05
Prob > F	=	0.0000
R-squared	=	0.0990
Adj R-squared	=	0.0850
Root MSE	=	1.676

haz	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
ex_r	.0020303	.0013157	1.54	0.123	-.0005508	.0046114
safewater	.0192869	.1093072	0.18	0.860	-.1951534	.2337272
improvedtoilets_vs1	.2902912	.1019296	2.85	0.004	.0903243	.4902581
advice2	.0962266	.1068725	0.90	0.368	-.1134372	.3058905
landholdings	.0405974	.0379195	1.07	0.285	-.0337936	.1149885
hhh_mature	-.0140334	.1038584	-0.14	0.893	-.2177842	.1897175
hhh_female	.0317556	.1871575	0.17	0.865	-.3354125	.3989236
province						
Central Province	.5943186	.2475298	2.40	0.016	.1087114	1.079926
Chimbu (Simbu) Province	.4455552	.2231952	2.00	0.046	.007688	.8834225
East New Britain Province	.9316815	.24666	3.78	0.000	.4477806	1.415582
East Sepik Province	.3376934	.2644979	1.28	0.202	-.1812021	.8565889
Eastern Highlands Province	.7891199	.2534324	3.11	0.002	.2919328	1.286307
Jiwaka Province	.5610519	.2476679	2.27	0.024	.0751739	1.04693
Madang Province	.7052731	.2313612	3.05	0.002	.2513857	1.15916
Morobe Province	-.4243679	.2527177	-1.68	0.093	-.9201529	.0714171
Western Province	1.267225	.1785[illegible]2	7.10	0.000	.9168979	1.617553
Western Highlands	.5562394	.21875[illegible]9	2.54	0.011	.1270851	.9853938
Gulf Province	1.293717	.2343655	5.52	0.000	.833936	1.753499
Oro Province	1.167069	.2215[illegible]1	5.27	0.000	.7324708	1.601668
Milne Bay Province	1.148766	.2669[illegible]3	4.30	0.000	.6249753	1.672557

- *East New Britain, Western, Gulf, Oro and Milne Bay province are significant (when compared to ARoB)*

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female i.province*

. regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female i.province

Source	SS	df	MS
Model	396.27377	20	19.8136885
Residual	3606.83887	1,284	2.80906454
Total	4003.11264	1,304	3.06987166

Number of obs	=	1,305
F(20, 1284)	=	7.05
Prob > F	=	0.0000
R-squared	=	0.0990
Adj R-squared	=	0.0850
Root MSE	=	1.676

haz	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
ex_r	.0020303	.0013157	1.54	0.123	-.0005508	.0046114
safewater	.0192869	.1093072	0.18	0.860	-.1951534	.2337272
improvedtoilets_vs1	.2902912	.1019296	2.85	0.004	.0903243	.4902581
advice2	.0962266	.1068725	0.90	0.368	-.1134372	.3058905
landholdings	.0405974	.0379195	1.07	0.285	-.0337936	.1149885
hhh_mature	-.0140334	.1038584	-0.14	0.893	-.2177842	.1897175
hhh_female	.0317556	.1871575	0.17	0.865	-.3354125	.3989236
province						
Central Province	.5943186	.2475298	2.40	0.016	.1087114	1.079926
Chimbu (Simbu) Province	.4455552	.2231952	2.00	0.046	.007688	.8834225
East New Britain Province	.9316815	.24666	3.78	0.000	.4477806	1.415582
East Sepik Province	.3376934	.2644979	1.28	0.202	-.1812021	.8565889
Eastern Highlands Province	.7891199	.2534324	3.11	0.002	.2919328	1.286307
Jiwaka Province	.5610519	.2476679	2.27	0.024	.0751739	1.04693
Madang Province	.7052731	.2313612	3.05	0.002	.2513857	1.15916
Morobe Province	-.4243679	.2527177	-1.68	0.093	-.9201529	.0714171
Western Province	1.267225	.1785732	7.10	0.000	.9168979	1.617553
Western Highlands	.5562394	.2187539	2.54	0.011	.1270851	.9853938
Gulf Province	1.293717	.2343655	5.52	0.000	.833936	1.753499
Oro Province	1.167069	.2215291	5.27	0.000	.7324708	1.601668
Milne Bay Province	1.148766	.2669933	4.30	0.000	.6249753	1.672557

- *This means that holding the effect of these other IV constant, children under 5 years in ENB Province have an average HAZ-score that is 0.931 higher than haz score of children under 5 years in ARoB.*

4. Ordinary least-squares (OLS) linear regression

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female i.province*

```
. regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female i.province
```

Source	SS	df	MS
Model	396.27377	20	19.8136885
Residual	3606.83887	1,284	2.80906454
Total	4003.11264	1,304	3.06987166

Number of obs	=	1,305
F(20, 1284)	=	7.05
Prob > F	=	0.0000
R-squared	=	0.0990
Adj R-squared	=	0.0850
Root MSE	=	1.676

haz	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
ex_r	.0020303	.0013157	1.54	0.123	-.0005508	.0046114
safewater	.0192869	.1093072	0.18	0.860	-.1951534	.2337272
improvedtoilets_vs1	.2902912	.1019296	2.85	0.004	.0903243	.4902581
advice2	.0962266	.1068725	0.90	0.368	-.1134372	.3058905
landholdings	.0405974	.0379195	1.07	0.285	-.0337936	.1149885
hhh_mature	-.0140334	.1038584	-0.14	0.893	-.2177842	.1897175
hhh_female	.0317556	.1871575	0.17	0.865	-.3354125	.3989236
province						
Central Province	.5943186	.2475298	2.40	0.016	.1087114	1.079926
Chimbu (Simbu) Province	.4455552	.2231952	2.00	0.046	.007688	.8834225
East New Britain Province	.9316815	.24666	3.78	0.000	.4477806	1.415582
East Sepik Province	.3376934	.2644979	1.28	0.202	-.1812021	.8565889
Eastern Highlands Province	.7891199	.2534324	3.11	0.002	.2919328	1.286307
Jiwaka Province	.5610519	.2476679	2.27	0.024	.0751739	1.04693
Madang Province	.7052731	.2313612	3.05	0.002	.2513857	1.15916
Morobe Province	-.4243679	.2527177	-1.68	0.093	-.9201529	.0714171
Western Province	1.267225	.17857[illegible]2	7.10	0.000	.9168979	1.617553
Western Highlands	.5562394	.2187539	2.54	0.011	.1270851	.9853938
Gulf Province	1.293717	.2343655	5.52	0.000	.833936	1.753499
Oro Province	1.167069	.22152[illegible]1	5.27	0.000	.7324708	1.601668
Milne Bay Province	1.148766	.2669933	4.30	0.000	.6249753	1.672557

- *Similarly, holding these other variables constant, children under 5 years in Gulf province have an average HAZ-score that is 1.29 higher than children under 5 years in ARoB, children under 5 years in Milne Bay province have average HAZ-score that is 1.14 higher than children under 5 years in ARoB and children under 5 years in Oro and Western province have a HAZ-score that is 1.16 and 1.27 higher than children under 5 years in ARoB respectively.*

4. Ordinary least-squares (OLS) linear regression

- Let's run one more regression, with more IVs

4. Ordinary least-squares (OLS) linear regression

- Let's run one more regression, with more IVs

- *regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female i.province*

4. Ordinary least-squares (OLS) linear regression

```
. regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female i.province

      Source |       SS           df       MS      Number of obs   =     1,305
-------------+----------------------------------   F(20, 1284)     =      7.05
       Model |  396.27377        20  19.8136885   Prob > F        =    0.0000
    Residual |  3606.83887     1,284  2.80906454   R-squared       =    0.0990
-------------+----------------------------------   Adj R-squared   =    0.0850
       Total |  4003.11264     1,304  3.06987166   Root MSE        =     1.676

------------------------------------------------------------------------------------------
                      haz | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
--------------------------+---------------------------------------------------------------
                     ex_r |   .0020303   .0013157     1.54   0.123    -.0005508    .0046114
                safewater |   .0192869   .1093072     0.18   0.860    -.1951534    .2337272
      improvedtoilets_vs1 |   .2902912   .1019296     2.85   0.004     .0903243    .4902581
                  advice2 |   .0962266   .1068725     0.90   0.368    -.1134372    .3058905
             landholdings |   .0405974   .0379195     1.07   0.285    -.0337936    .1149885
               hhh_mature |  -.0140334   .1038584    -0.14   0.893    -.2177842    .1897175
               hhh_female |   .0317556   .1871575     0.17   0.865    -.3354125    .3989236
                          |
                 province |
        Central Province  |   .5943186   .2475298     2.40   0.016     .1087114    1.079926
  Chimbu (Simbu) Province |   .4455552   .2231952     2.00   0.046      .007688    .8834225
 East New Britain Province |   .9316815     .24666     3.78   0.000     .4477806    1.415582
      East Sepik Province |   .3376934   .2644979     1.28   0.202    -.1812021    .8565889
Eastern Highlands Province |   .7891199   .2534324     3.11   0.002     .2919328    1.286307
          Jiwaka Province |   .5610519   .2476679     2.27   0.024     .0751739     1.04693
          Madang Province |   .7052731   .2313612     3.05   0.002     .2513857     1.15916
          Morobe Province |  -.4243679   .2527177    -1.68   0.093    -.9201529    .0714171
         Western Province |   1.267225   .1785732     7.10   0.000     .9168979    1.617553
        Western Highlands |   .5562394   .2187539     2.54   0.011     .1270851    .9853938
            Gulf Province |   1.293717   .2343655     5.52   0.000      .833936    1.753499
             Oro Province |   1.167069   .2215291     5.27   0.000     .7324708    1.601668
       Milne Bay Province |   1.148766   .2669933     4.30   0.000     .6249753    1.672557
```

4. Ordinary least-squares (OLS) linear regression

```
. regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female i.province

      Source |       SS           df       MS      Number of obs   =     1,305
-------------+----------------------------------   F(20, 1284)     =      7.05
       Model |  396.27377        20  19.8136885   Prob > F        =    0.0000
    Residual |  3606.83887     1,284  2.80906454   R-squared       =    0.0990
-------------+----------------------------------   Adj R-squared   =    0.0850
       Total |  4003.11264     1,304  3.06987166   Root MSE        =     1.676

----------------------------------------------------------------------------------------------
                         haz | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-----------------------------+----------------------------------------------------------------
                        ex_r |   .0020303   .0013157     1.54   0.123    -.0005508    .0046114
                   safewater |   .0192869   .1093072     0.18   0.860    -.1951534    .2337272
         improvedtoilets_vs1 |   .2902912   .1019296     2.85   0.004     .0903243    .4902581
                     advice2 |   .0962266   .1068725     0.90   0.368    -.1134372    .3058905
                landholdings |   .0405974   .0379195     1.07   0.285    -.0337936    .1149885
                  hhh_mature |  -.0140334   .1038584    -0.14   0.893    -.2177842    .1897175
                  hhh_female |   .0317556   .1871575     0.17   0.865    -.3354125    .3989236
                             |
                    province |
           Central Province  |   .5943186   .2475298     2.40   0.016     .1087114    1.079926
     Chimbu (Simbu) Province  |   .4455552   .2231952     2.00   0.046      .007688    .8834225
    East New Britain Province |   .9316815     .24666     3.78   0.000     .4477806    1.415582
         East Sepik Province |   .3376934   .2644979     1.28   0.202    -.1812021    .8565889
 Eastern Highlands Province  |   .7891199   .2534324     3.11   0.002     .2919328    1.286307
            Jiwaka Province  |   .5610519   .2476679     2.27   0.024     .0751739     1.04693
            Madang Province  |   .7052731   .2313612     3.05   0.002     .2513857     1.15916
            Morobe Province  |  -.4243679   .2527177    -1.68   0.093    -.9201529    .0714171
           Western Province  |   1.267225   .1785732     7.10   0.000     .9168979    1.617553
          Western Highlands  |   .5562394   .2187539     2.54   0.011     .1270851    .9853938
              Gulf Province  |   1.293717   .2343655     5.52   0.000      .833936    1.753499
               Oro Province  |   1.167069   .2215291     5.27   0.000     .7324708    1.601668
         Milne Bay Province  |   1.148766   .2669933     4.30   0.000     .6249753    1.672557
```

- What's significant?

4. Ordinary least-squares (OLS) linear regression

. regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female i.province

Source	SS	df	MS
Model	396.27377	20	19.8136885
Residual	3606.83887	1,284	2.80906454
Total	4003.11264	1,304	3.06987166

Number of obs	=	1,305
F(20, 1284)	=	7.05
Prob > F	=	0.0000
R-squared	=	0.0990
Adj R-squared	=	0.0850
Root MSE	=	1.676

haz	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
ex_r	.0020303	.0013157	1.54	0.123	-.0005508	.0046114
safewater	.0192869	.1093072	0.18	0.860	-.1951534	.2337272
improvedtoilets_vs1	.2902912	.1019296	2.85	0.004	.0903243	.4902581
advice2	.0962266	.1068725	0.90	0.368	-.1134372	.3058905
landholdings	.0405974	.0379195	1.07	0.285	-.0337936	.1149885
hhh_mature	-.0140334	.1038584	-0.14	0.893	-.2177842	.1897175
hhh_female	.0317556	.1871575	0.17	0.865	-.3354125	.3989236
province						
Central Province	.5943186	.2475298	2.40	0.016	.1087114	1.079926
Chimbu (Simbu) Province	.4455552	.2231952	2.00	0.046	.007688	.8834225
East New Britain Province	.9316815	.24666	3.78	0.000	.4477806	1.415582
East Sepik Province	.3376934	.2644979	1.28	0.202	-.1812021	.8565889
Eastern Highlands Province	.7891199	.2534324	3.11	0.002	.2919328	1.286307
Jiwaka Province	.5610519	.2476679	2.27	0.024	.0751739	1.04693
Madang Province	.7052731	.2313612	3.05	0.002	.2513857	1.15916
Morobe Province	-.4243679	.2527177	-1.68	0.093	-.9201529	.0714171
Western Province	1.267225	.1785732	7.10	0.000	.9168979	1.617553
Western Highlands	.5562394	.2187539	2.54	0.011	.1270851	.9853938
Gulf Province	1.293717	.2343655	5.52	0.000	.833936	1.753499
Oro Province	1.167069	.2215291	5.27	0.000	.7324708	1.601668
Milne Bay Province	1.148766	.2669933	4.30	0.000	.6249753	1.672557

- What's significant?
- East New Britain, Milne Bay, Oro, Gulf, Western province have p-values=0.000.
- This means that holding the effect of these other variables constant, children under 5 years in Western Province have an average HAZ-score that is 1.30 higher than children under 5 years in ARoB.
- Similarly, holding these other variables constant, children under 5 years in Gulf province have an average HAZ-score that is 1.29 higher than children under 5 years in ARoB, children under 5 years in Milne Bay province have average HAZ-score that is 1.14 higher than children under 5 years in ARoB and children under 5 years in Oro province have a HAZ-score that is 1.16 higher than children under 5 years in ARoB.