



Department of Foreign Affairs and Trade





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

1. <u>Review of Lessons 1 and 2</u>



1. <u>Review of Lessons 1 and 2</u>

2. <u>Correlations</u>

1. <u>Review of Lessons 1 and 2</u>

2. <u>Correlations</u>

3. <u>T-Tests</u>

1. <u>Review of Lessons 1 and 2</u>

2. <u>Correlations</u>

3. <u>T-Tests</u>

4. Ordinary least-squares (OLS) linear regressions

1. <u>Review of Lessons 1 and 2</u>

2. <u>Correlations</u>

3. <u>T-Tests</u>

4. Ordinary least-squares (OLS) linear regressions

5. <u>Child Height-for-Age z-score (HAZ-score) presentation</u>

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

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

- 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*

- 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?

- 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

- 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?

- 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

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

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

How many households have a female household head?

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

How many households have a female household head?
tab hhh_female -

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%)

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?

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 -

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)

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

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)

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

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

Copy child information from Roster page

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

	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

Сор	Copy child information from Roster page			5years of age						
<	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_man ner_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

Сор	Copy child information from Roster page			5years of age							
\leq	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_man ner_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									
				·	·	·					
				·		ght and					
					Heigh						
					under	r 5 years					

- 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.

- 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

- 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?

- 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

- 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?

- 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

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

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

hist haz

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



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

▶ hist haz

What can we learn from this figure?



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)



We have some continuous data in this dataset

Examples: landholdings and household size (hhsize)



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)

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

▶ Many times there are notable *outliers* in raw, continuous data

Outlier - an observation very different from all other observations

- Outlier an observation very different from all other observations
- What's a good way to check landholdings for outliers?

- Outlier an observation very different from all other observations
- What's a good way to check landholdings for outliers?
 sum landholdings, det

▶ 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

. sun	n landholdings,	det		
	Size of	all landholdings	in hectares	
	Percentiles	Smallest		
18	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

▶ 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?





The median is 1.05, but the max is 19.55!

- Outlier an observation very different from all other observations
- What's a good way to check landholdings for outliers?
 - sum landholdings, det
 - hist landholdings

- Outlier an observation very different from all other observations
- What's a good way to check landholdings for outliers?
 - sum landholdings, det
 - hist landholdings



- Outlier an observation very different from all other observations
- What's a good way to check landholdings for outliers?
 - sum landholdings, det
 - hist landholdings



Everyone has a different preference for handling outliers

Everyone has a different preference for handling outliers

Let's change all values greater than the 99th percentile to the median

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

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

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 a	all landholding	s 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	\supset	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

- 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, a	de
---------------------	----

- ▶ *Median: 1.05*
- ▶ 99th percentile: 8.0824

	. sum	landholdings,	det		
		Size of	all landholdings	in hectares	
		Percentiles	Smallest		
	1%	0	0		
1	5%	.0162	0		
T	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

- 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"

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

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

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

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

Replace landholding=1.05 if landholdings>8.0824

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?

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)

Now let's see what the new distribution looks like

Now let's see what the new distribution looks like

hist landholdings

Now let's see what the new distribution looks like

hist landholdings



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!



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

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

<u>Code:</u> pwcorr haz ex_r

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

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

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

pwcorr haz ex_r

. pwcorr haz ex_r			
	haz	ex_r	
haz ex_r	1.0000 0.1194	1.0000	
pwcorr haz ex_r

What is the correlation coefficient?

•	pwcorr	haz	ex_r	
			haz	ex_r
	e	haz ex_r	1.0000 0.1194	1.0000

pwcorr haz ex_r

What is the correlation coefficient? 0.1194



pwcorr haz ex_r

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

. pwcorr haz e	x_r
	haz ex_r
haz ex r	1.0000

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	
	e	ex_r	0	.1194	>1.0000

- 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 e	x_r
	haz ex_r
haz ex r	1.0000

- 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 e	x_r
	haz ex_r
haz ex_r	1.0000 0.1194 1.0000

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

. pwcorr haz ex_r				
	haz ex_r			
haz ex_r	1.0000 0.1194 1.0000			

- 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 e	ex_r
	haz ex_r
haz ex_r	1.0000 0.1194 1.0000

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

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

<u>Code:</u> pwcorr haz ex_r, sig

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

\sim Coue. procord huz ex_1, sig		Code:	pwcorr	haz	ex_	_r,	sig	
------------------------------------	--	-------	--------	-----	-----	-----	-----	--

pwcorr haz ex_r, sig				
	haz	ex_r		
haz	1.0000			
ex_r	0. 1194 0.0000	1.0000		

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 e	ex_r, sig	
	haz	ex_r
haz	1.0000	
ex_r	0.1194 0.0000	1.0000

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



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?



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



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	

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 e	ex_r, sig
	haz ex_r
haz	1.0000
ex_r	0.1194 1.0000

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

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

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

<u>Code:</u> scatter haz ex_r

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

<u>Code:</u> scatter haz ex_r



- 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



<u>Code:</u> twoway (scatter haz ex_r) (lfit haz ex_r)

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



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

<u>Code:</u> ttest haz, by(meat_poul_fish_C)

<u>Code:</u> ttest haz, by(meat_poul_fish_C)

. ttest ha	az, by(mea	at_poul_fish_	C)			
Two-sample	e t test wi	ith equal v ar	iances			
Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
No Yes	224 374	-1.950223 -1.637834	.0990268 .0827858	1.482097 1.601001	-2.145371 -1.80062	-1.755075 -1.475049
combined	598	-1.754849	.0639408	1.56361	-1.880426	-1.629273
diff		312389	.1315951		5708356	0539424
diff = Ho: diff =	= mean (No) = 0	- mean(Yes)		degrees	t : of freedom :	= -2.3739 = 596
Ha: d: Pr(T < t)	iff < 0) = 0.0090	Pr(Ha: diff != T > t) = (0 .0179	Ha: d Pr(T > t	iff > 0) = 0.9910

Code: ttest haz, by(meat_poul_fish_C)

. ttest ha	az, by(mea	at_poul_fish_	C)			
Two-sample	e t test wi	ith equal v ar	iances			
Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
No Yes	224 374	-1.950223 -1.637834	.0990268 .0827858	1.482097 1.601001	-2.145371 -1.80062	-1.755075 -1.475049
combined	598	-1.754849	.0639408	1.56361	-1.880426	-1.629273
diff		312389	.1315951		5708356	0539424
diff = Ho: diff =	= mean(No) = 0	- mean(Yes)		degrees	t : of freedom :	= -2.3739 = 596
Ha: di Pr(T < t)	lff < 0 = 0.0090	Pr(Ha: diff != T > t) = (0 0.0179	Ha: d Pr(T > t	iff > 0) = 0.9910

What does this output tell us?

Code: ttest haz, by(meat_poul_fish_C)

. ttest ha	az, by(mea	at_poul_fish_	C)			
Two-sample	e t test w	ith equal v ar	iances			
Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
No Yes	224 374	-1.950223 -1.637834	.0990268 .0827858	1.482097 1.601001	-2.145371 -1.80062	-1.755075 -1.475049
combined	598	-1.754849	.0639408	1.56361	-1.880426	-1.629273
diff		312389	.1315951		5708356	0539424
diff = Ho: diff =	= mean(No) = 0	- mean(Yes)		degrees	t : of freedom :	= -2.3739 = 596
Ha: d: Pr(T < t)	iff < 0) = 0.0090	Pr(Ha: diff != T > t) = (0 0.0179	Ha: d Pr(T > t	iff > 0) = 0.9910

► The number of observations in each group

Code: ttest haz, by(meat_poul_fish_C)

. ttest ha	az, by(me	at_poul_fish_	C)			
Two-sample	e t test w	ith equal v ar	iances			
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: di	iff < 0		Ha: diff !=	0	Ha: d	iff > 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 <u>not</u> eat meat, poultry or fish in the last 24 hours.

Code: ttest haz, by(meat_poul_fish_C)

. ttest ha	az, by(mea	at_poul_fish_C	:)			
Two-sample	e t test w	ith equal vari	lances			
Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
No Yes	224 374	-1.950223 -1.637834	.0990268 .0827858	1.482097 1.601001	-2.145371 -1.80062	-1.755075 -1.475049
combined	598	-1.754849	.0639408	1.56361	-1.880426	-1.629273
diff		312389	.1315951		5708356	0539424
diff = Ho: diff =	= mean (No) = 0	- mean(Yes)		degrees	t = of freedom =	= -2.3739 = 596
Ha: d: Pr(T < t)	iff < 0) = 0.0090	Pr(1	Ha: diff != ? > t) = (0 .0179	Ha: d: Pr(T > t)	iff > 0) = 0.9910



Code: ttest haz, by(meat_poul_fish_C)

. ttest ha	az, by(mea	at_poul_fish_C	:)			
Two-sample	e t test w	ith equal vari	lances			
Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
No Yes	224 374	-1.950223 -1.637834	.0990268 .0827858	1.482097 1.601001	-2.145371 -1.80062	-1.755075 -1.475049
combined	598	-1.754849	.0639408	1.56361	-1.880426	-1.629273
diff		312389	.1315951		5708356	0539424
diff = Ho: diff =	= mean (No) = 0	- mean(Yes)		degrees	t = of freedom =	= -2.3739 = 596
Ha: d: Pr(T < t)	iff < 0) = 0.0090	Pr(1	Ha: diff != ? > t) = (0 .0179	Ha: d: Pr(T > t)	iff > 0) = 0.9910

Which group has a higher average haz-score?

Code: ttest haz, by(meat_poul_fish_C)

. ttest ha	az, by(mea	at_poul_fish_C	2)			
Two-sample	e t test wi	ith equal v ari	lances			
Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
No Yes	224 374	-1.950223 -1.637834	.0990268 .0827858	1.482097 1.601001	-2.145371 -1.80062	-1.755075 -1.475049
combined	598	-1.754849	.0639408	1.56361	-1.880426	-1.629273
diff		312389	.1315951		5708356	0539424
diff = Ho: diff =	= mean (No) = 0	- mean(Yes)		degrees	t : of freedom :	= -2.3739 = 596
Ha: di Pr(T < t)	iff < 0) = 0.0090	Pr(1	Ha: diff != ? > t) = (0 0.0179	Ha: d Pr(T > t	iff > 0) = 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)

Code: ttest haz, by(meat_poul_fish_C)

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

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

Code: ttest haz, by(meat_poul_fish_C)

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
No Yes	224 374	-1.950223 -1.637834	.0990268 .0827858	1.482097 1.601001	-2.145371 -1.80062	-1.755075 -1.475049
combined	598	-1.754849	.0639408	1.56361	-1.880426	-1.629273
diff		312389	.1315951		5708356	0539424
diff : Ho: diff :	= mean(No) = 0	- mean(Yes)		degrees	t : of freedom :	= -2.373 = 59

Do the two confidence intervals overlap?

Code: ttest haz, by(meat_poul_fish_C)

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

Do the two confidence intervals overlap?

▶ No - the max for "no" is -1.755 and the min for "yes" is -1.801
<u>Code</u>: ttest haz, by(meat_poul_fish_C)

. ttest ha	ttest haz, by(meat_poul_fish_C)									
Two-sample	e t test w:	ith equal v ar	iances							
Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]				
No Yes	224 374	-1.950223 -1.637834	.0990268 .0827858	1.482097 1.601001	-2.145371 -1.80062	-1.755075 -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: d: Pr(T < t)	iff < 0) = 0.0090	Pr(Ha: diff != T > t) = (0 0.0179	Ha: d Pr(T > t	iff > 0) = 0.9910				

► T-statistic and the degrees of freedom

Code: ttest haz, by(meat_poul_fish_C)

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

► 3 different p-values

Code: ttest haz, by(meat_poul_fish_C)

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

P-value for whether the mean of Group 1 is <u>less than</u> the mean of Group 2

Code: ttest haz, by(meat_poul_fish_C)

. ttest ha	. ttest haz, by(meat_poul_fish_C)										
Two-sample	e t test w	ith equa	l varia	ances							
Group	Obs	М	lean	Std.	Err.	Std.	Dev.	[95%	Conf.	Inter	val]
No Yes	224 374	-1.950 -1.637	223 834	.0990 .0827	268 858	1.482 1.601	097 001	-2.14	5371 0062	-1.75 -1.47	55075 75049
combined	598	-1.754	849	.0639	408	1.56	361	-1.88	0426	-1.62	29273
diff		312	389	.1315	951			570	8356	053	39424
diff = Ho: diff =	= mean(No) = 0	- mean(Yes)			de	grees	of fre	t : eedom :	= -2 . =	. 3739 596
Ha: d: Pr(T < t)	iff < 0) = 0.0090		H Pr(T	Ha: di > t	ff !=) = 0	0 . 0179		Pr	Ha: d: (T > t)	iff >) = 0 .	0 9910

P-value for whether the mean of Group 1 is <u>not equal</u> to the mean of Group 2

Code: ttest haz, by(meat_poul_fish_C)

. ttest haz, by(meat_poul_fish_C)								
Two-sample	e t test wi	ith equal v ar	iances					
Group	Obs	Mean	Std. Err.	Std. Dev.	[9	95% Conf.	Interval]	
No Yes	224 374	-1.950223 -1.637834	.0990268 .0827858	1.482097 1.601001	-2. -1	145371 80062	-1.755075 -1.475049	
combined	598	-1.754849	.0639408	1.56361	-1.	880426	-1.629273	
diff		312389	.1315951		5	708356	0539424	
diff = Ho: diff =	= mean (No) = 0	- mean(Yes)		degrees	of	t : freedom :	= -2.3739 = 596	
Ha: diff < 0Ha: diff != 0Ha: diff > 0 $Pr(T < t) = 0.0090$ $Pr(T > t) = 0.0179$ $Pr(T > t) = 0.991$							iff > 0) = 0.9910	

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

Code: ttest haz, by(meat_poul_fish_C)

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

What is the p-value for whether the mean of Group 1 is <u>less</u> <u>than</u> the mean of Group 2

Code: ttest haz, by(meat_poul_fish_C)

. ttest ha	az, by(mea	t_poul_fish_	C)			
Two-sample	e t test wi	th equal v ar	iances			
Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
No Yes	224 374	-1.950223 -1.637834	.0990268 .0827858	1.482097 1.601001	-2.145371 -1.80062	-1.755075 -1.475049
combined	598	-1.754849	.0639408	1.56361	-1.880426	-1.629273
diff		312389	.1315951		5708356	0539424
diff = Ho: diff =	= mean(No) = 0	- mean(Yes)		degrees	t : of freedom :	= -2.3739 = 596
Ha: diff < 0						

What is the p-value for whether the mean of Group 1 is <u>less</u> <u>than</u> the mean of Group 2

▶ *P-value is* 0.0090

Code: ttest haz, by(meat_poul_fish_C)

. ttest ha	<pre>ttest haz, by(meat_poul_fish_C)</pre>								
Two-sample	e t test wi	th equal var	iances						
Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]			
No Yes	224 374	-1.950223 -1.637834	.0990268 .0827858	1.482097 1.601001	-2.145371 -1.80062	-1.755075 -1.475049			
combined	598	-1.754849	.0639408	1.56361	-1.880426	-1.629273			
diff		312389	.1315951		5708356	0539424			
diff = Ho: diff =	= mean(No) = 0	- mean(Yes)		degrees	t of freedom	= -2.3739 = 596			
Ha: d: Pr(T < t)	iff < 0) = 0.0090	Pr(Ha: diff != T > t) = (0 0.0179	Ha: d Pr(T > t	iff > 0) = 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.

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.

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.



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)

ttest haz, by(hhh_mature)

ttest haz, by(hhh_mature)

. ttest ha	az, by(hhl							
Two-sample	e t test wi	ith equal var	riances					
Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]		
Youth he Mature h	411 894	-1.435815 -1.317058	.085222 .0589603	1.727717 1.762903	-1.603342 -1.432775	-1.268289 -1.201341		
combined	1,305	-1.35446	.0485015	1.752105	-1.449609	-1.25931		
diff		1187569	.1044063		3235798	.086066		
diff = Ho: diff =	diff = mean(Youth he) - mean(Mature h) $t = -1.1374$ Ho: diff = 0 degrees of freedom = 1303							
Ha: di Pr(T < t)	iff < 0 = 0.1278	Pr(Ha: diff != T > t) = (0 0.2556	Ha: d Pr(T > t	iff > 0) = 0.8722		

ttest haz, by(hhh_mature)

. ttest ha	az, by(hhl						
Two-sample	e t test wi	ith equal var	iances				
Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]	
Youth he Mature h	411 894	-1.435815 -1.317058	.085222 .0589603	1.727717 1.762903	-1.603342 -1.432775	-1.268289 -1.201341	
combined	1,305	-1.35446	.0485015	1.752105	-1.449609	-1.25931	
diff		1187569	.1044063		3235798	.086066	
diff = Ho: diff =	diff = mean(Youth he) - mean(Mature h) t = -1.1374 Ho: diff = 0 degrees of freedom = 1303						
Ha: di Pr(T < t)	iff < 0 = 0.1278	Pr(Ha: diff != T > t) = (0 0.2556	Ha: d Pr(T > t	iff > 0) = 0.8722	

Which mean is higher?

ttest haz, by(hhh_mature)

. ttest ha	az, by(hhh	_mature)					
Two-sample	e t test wi	th equal vari	ances				
Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]	
Youth he Mature h	411 894	-1.435815 -1.317058	.085222 .0589603	1.727717 1.762903	-1.603342 -1.432775	-1.268289 -1.201341	
combined	1,305	-1.35446	.0485015	1.752105	-1.449609	-1.25931	
diff		1187569	.1044063		3235798	.086066	
diff = Ho: diff =	diff = mean(Youth he) - mean(Mature h) $t = -1.1374$ Ho: diff = 0 degrees of freedom = 1303						
Ha: di Pr(T < t)	lff < 0 = 0.1278	Pr(T	Ha: diff != > t) = (0 .2556	Ha: d Pr(T > t	iff > 0) = 0.8722	

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

ttest haz, by(hhh_mature)

. ttest ha	az, by(hhł	_mature)				
Two-sample	e t test wi	ith equal var	iances			
Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
Youth he Mature h	411 894	-1.435815 -1.317058	.085222 .0589603	1.727717 1.762903	-1.603342 -1.432775	-1.268289 -1.201341
combined	1,305	-1.35446	.0485015	1.752105	-1.449609	-1.25931
diff		1187569	.1044063		3235798	.086066
diff = Ho: diff =	= mean(Yout = 0	ch he) - mean	(Mature h)	degrees	t ; of freedom	= -1.1374 = 1303
Ha: di Pr(T < t)	iff < 0) = 0.1278	Pr(Ha: diff != T > t) = (0 0.2556	Ha: d Pr(T > t	liff > 0 () = 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?

ttest haz, by(hhh_mature)

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval
Youth he Mature h	411 894	-1.435815 -1.317058	.085222 .0589603	1.727717 1.762903	-1.603342 -1.432775	-1.268289 -1.201343
combined	1,305	-1.35446	.0485015	1.752105	-1.449609	-1.2593
diff		1187569	.1044063		3235798	.08606
diff diff =	= mean(You)	1187569 th he) - mean	.1044063	dograa	3235798	.086

- 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

Now we are going to move on to our final analysis - an 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)

- 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

regress haz ex_r

regress haz ex_r

. regress haz	ex_r						
Source	SS	df	MS	Numb	per of obs	=	1,305
Model Residual	57.052925 3946.05972	1 1,303	57.052925 3.02844184	- F(1, 5 Prok 4 R-sc	1303) > F quared	= =	0.0000
Total	4003.11264	1,304	3.06987166	- Adj 6 Root	R-squared MSE	=	0.0135 1.7402
haz	Coef.	Std. Err.	t	P> t	[95% Cor	nf.	Interval]
ex_r _cons	.0052426 -1.650508	.0012079 .083504	4.34 -19.77	0.000	.002873 -1.814325	3 5	.0076121 -1.486691

regress haz ex_r

. regress haz	ex_r						
Source	SS	df	MS	Numk	per of obs	=	1,305
				- F(1,	1303)	=	18.84
Model	57.052925	1	57.05292	25 Prob > F		=	0.0000
Residual	3946.05972	1,303	3.0284418	.844184 R-squared		=	0.0143
,				- Adj	Adj R-squared		0.0135
Total	4003.11264	1,304	3.0698716	6 Root	MSE	=	1.7402
haz	Coef.	Std. Err.	t	P> t	[95% Co	nf.	Interval]
ex r	.0052426	.0012079	4.34	0.000	.00287	3	.0076121
_cons	-1.650508	.083504	-19.77	0.000	-1.81432	5	-1.486691

So what does this output show us?

regress haz ex_r

. regress haz	ex_r							
Source	SS	df	MS		Numbe	r 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	3.0698716	3.06987166		Root MSE		1.7402
haz	Coef.	Std. Err.	t	P>	· t	[95% Cor	nf.	Interval]
ex r	.0052426	.0012079	4.34	0.	000	.002873	3	.0076121
_cons	-1.650508	.083504	-19.77	0.	000	-1.814325	5	-1.486691

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

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	5 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	P> t [95% Cc	onf.	Interval]
ex_r	.0052426	.0012079	4.34 0	.000 .00287	3	.0076121
_cons	-1.650508	.083504	-19.77 0	.000 -1.81432	25	-1.486691

Overall model fit - number of observations, F-statistic, Rsquared, etc.

regress haz ex_r

. regress haz	ex_r							
Source	SS	df	MS	Numb	er of obs	=	1,305	
				— F(1, 1303)		=	18.84	
Model	57.052925	1	57.052925	Prob	Prob > F		0.0000	
Residual	3946.05972	1,303	3.02844184	R-sq	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% Co	nf.	Interval]	
ex r	.0052426	.0012079	4.34	0.000	.002873	3	.0076121	
_cons	-1.650508	.083504	-19.77	0.000	-1.81432	5	-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)

regress haz ex_r

. regress haz	ex_r						
Source	SS	df	MS	Nur	Number of obs		1,305
				— F(1	- F(1, 1303)		18.84
Model	57.052925	1	57.0529	925 Pro	Prob > F		0.0000
Residual	3946.05972	1,303	3.028441	184 R-s	R-squared		0.0143
				— Adj	Adj R-squared		0.0135
Total	4003.11264	1,304	3.069871	L66 Roc	Root MSE		1.7402
haz	Coef.	Std. Err.	t	P> t	[95% Conf. Inte		interval]
ex r	.0052426	.0012079	4.34	0.000	.00287	3	.0076121
cons	-1.650508	.083504	-19.77	0.000	-1.81432	5 -	1.486691

Does this p-value look familiar at all?

regress haz ex_r

. regress haz	ex_r							
Source	SS	df	MS	Nur	mber of obs	=	1,305	
			——— F(1, 1303)		— F(1, 1303)		18.84	
Model	57.052925	1	57.0529	925 Pro	25 Prob > F		0.0000	
Residual	3946.05972	1,303	3.028441	184 R-:	4 R-squared		0.0143	
				Ad	Adj R-squared		0.0135	
Total	4003.11264	1,304	3.069871	L66 Roo	66 Root MSE		1.7402	
haz	Coef.	Std. Err.	t	P> t	t [95% Cor		Interval]	
ex_r	.0052426	.0012079	4.34	0.000	.002873	3	.0076121	
_cons	-1.650508	.083504	-19.77	0.000	-1.81432	5.	-1.486691	

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

regress haz ex_r

. regress haz	ex_r						
Source	SS	df	MS	Nun	ber of ob	s =	1,305
				— F(1	, 1303)	=	18.84
Model	57.052925	1	57.052925 P		925 Prob > F		0.0000
Residual	3946.05972	1,303	3.02844184 R-so		quared	=	0.0143
				— Adj	R-squared	= b	0.0135
Total	4003.11264	1,304	3.069871	.66 Roc	t MSE	=	1.7402
haz	Coef.	Std. Err.	t	P> t	[95% (Conf.	Interval]
ex_r	.0052426	.0012079	4.34	0.000	.0028	873	.0076121
_cons	-1.650508	.083504	-19.77	0.000	-1.8143	325	-1.486691



- 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

Now, let's try adding some more variables

► Now, let's try adding some more variables

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

▶ 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).

► 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).



- 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

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
- What types of variables are the IVs in this 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

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)

- 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

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

 regress haz 	ex_r s	afewater	improv	edtoilets_	vs1 advi	ce2 landho	ldings	hhh_m	ature hhh_fema
Source		SS	ď	f MS	Nu	mber of ob	s =	1,	305
					— F(7, 1297)	=	5	.75
Model	120	572361		7 17.2246	229 Pr	ob > F	=	0.0	900
Residual	3882	2.54028	1,29	7 2.99347	747 R-	squared	=	0.0	301
					— Ad	j R-square	d =	0.0	249
Total	4003	3.11264	1,30	4 3.06987	' 166 Ro	ot MSE	=	1.7	302
	haz	Coeffi	cient :	Std. err.	t	P> t	[95%	conf.	interval]
	ex_r	.004	6134	.0012492	3.69	0.000	.002	1627	.0070641
saf	ewater	.078	9408	.1091627	0.72	0.470	13	5214	.2930956
improvedtoile [.]	ts vs1	.360	4228	.100654	3.58	0.000	.162	9602	.5578853
a	dvice2	.190	7452	.0968147	1.97	0.049	.000	8146	.3806757
landho	ldings	06	5619	.0351863	-1.86	0.062	134	5472	.0034093
hhh i	mature	.041	2775	.1060595	0.39	0.697	166	7895	.2493445
hhh	female	.01	0044	.1911317	0.05	0.958	364	9172	.3850051
_	_cons	-1.80	0801	.1246846	-14.44	0.000	-2.04	5406	-1.556195

regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female

. regress haz	ex_r s	afewater	improve	dtoilets_	vs1	advice2	2 landho	ldings	hhh_m	ature hhh_1	Female
Source		SS	df	MS	5	Numbe	er of ob	s =	1,	305	
Model	120.	572361	7	17.2246	5229	F(7, Prob	1297) > F	=	5 0.0	.75 000	
Residual	3882	.54028	1,297	2.99347	7747	R-squ	uared	=	0.0	301	
Total	4003	.11264	1,304	3.06987	7166	Root	MSE		1.7	249 302	
	haz	Coeffi	cient S	td. err.		t F	?> t	[95%	conf.	interval]	
	ex_r	.004	6134 .	0012492	3	.69 @	000	.002	1627	.0070641	
safe	ewater	.078	9408 .	1091627	0	.72 @	9.470	13	5214	.2930956	
improvedtoilet	ts_vs1	.360	4228	.100654	3	.58 6	0.000	.162	9602	.5578853	
ad	dvice2	.190	7452 .	0968147	1	.97 6	0.049	.000	8146	.3806757	
landhol	ldings	06	5619 .	0351863	-1	.86 6	0.062	134	5472	.0034093	
hhh_n	nature	.041	2775 .	1060595	0	.39 6	0.697	166	7895	.2493445	
hhh_f	female	.01	0044 .	1911317	0	.05 0	9.958	364	9172	.3850051	
	_cons	-1.80	0801 .	1246846	-14	.44 @	000	-2.04	5406	-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 e	ex_r safewate	er improve	dtoilets_vs1	advice2 la	ndholdings	hhh_mature	hhh_femal
Source	SS	df	MS	Number o	fobs =	1,305	
Model	120.572361	7	17.2246229	F(7, 129 Prob > F	/) =	5.75 0.0000	
Residual	3882.54028	1,297	2.99347747	R-square Adi R-sa	d = uared =	0.0301 0.0249	
Total	4003.11264	1,304	3.06987166	Root MSE	=	1.7302	
	haz Coef	icient S	td. err.	t P> t	[95%	conf. inter	rval]
safew	ex_r .00 vater .01	946134 . 789408 .	0012492 1091627	3.69 0.00 0.72 0.47	0 .002 013	1627 .00 5214 .29	70641 30956
improvedtoilets adv	s_vs1 .30 /ice2 .19	04228 07452 .	.100654 0968147	3.58 0.00 1.97 0.04	0.1629 9.000	9602 .553 8146 .380	78853 06757
landhold	lings(65619 .	9351863 -	1.86 0.06	2134	6472 . 00	34093
hhh_ma hhh_fe	ature .04 emale .0	12775 . 10044 .	1060595 1911317	0.39 0.69 0.05 0.95	7166 8364	7895 .249 9172 .38	93445 50051
-	_cons -1.8	. 800801	1246846 -1	4.44 0.00	0 -2.04	5406 -1.5	56195

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

regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female

 regress 	haz	ex_r s	afewater	improve	edtoilets	_vs1	advice2	2 landho	ldings	hhh_m	ature hhh_ [.]	female
Sou	rce		SS	df	F M	S	Numbe	er of ob	s =	1,	305	
Мо	del	120.	572361	7	17.224	6229	F(7, Prob	1297) > F	=	5 0.0	.75 000	
Resid	ual	3882	.54028	1,297	2.9934	7747	R-squ Adi R	uared R-squared	=	0.0	301 249	
То	tal	4003	.11264	1,304	3.0698	7166	Root	MSE	=	1.7	302	
		haz	Coeffi	cient S	Std. err.		t F	•> t	[95%	conf.	interval]	
		ex_r	.004	6134 .	0012492	3	3.69 0	000.000	.002	1627	.0070641	
	safe	ewater	.078	9408 .	1091627	6	9.72 6	9.470	13	5214	.2930956	
improvedt	oilet	ts_vs1	.360	4228	.100654	3	3.58 6	0.000	.162	9602	.5578853	
	ac	dvice2	.190	7452 .	0968147	1	L.97 Ø	0.049	.000	8146	.3806757	
la	ndho]	ldings	06	5619 .	0351863	-1	L.86 @	0.062	134	6472	.0034093	
	hhh_n	nature	.041	2775 .	1060595	6	9.39 6	0.697	166	7895	.2493445	
	hhh_ 1	Female	.01	0044 .	1911317	6	9.05 6	9.958	364	9172	.3850051	
		_cons	-1.80	0801 .	1246846	-14	4.44 0	9.000	-2.04	5406	-1.556195	

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

regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female

. regress ha	z ex_r s	afewater	improved	toilets_v	vs1 adv:	ice2 landho	ldings	hhh_m	ature hhh_	female
Source		SS	df	MS	N	umber of ob	s =	1,	305	
Madal	120	572264	-	17 2246	— F	(7, 1297)	=	5	.75	
Model	120.	572361	/	17.22462	229 PI	"OD > F	=	0.0	000	
Residual	3882	.54028	1,297	2.993477	747 R	-squared	=	0.0	301	
					— A	dj R-square	d =	0.0	249	
Total	4003	.11264	1,304	3.069871	166 Ro	ot MSE	=	1.7	302	
	haz	Coeffic	ient St	d. err.	t	P> t	[95%	conf.	interval]	
	ex_r	.0046	134 .0	012492	3.69	0.000	.002	1627	.0070641	
sa	fewater	.0789	408 .1	.091627	0.72	0.470	13	5214	.2930956	
improvedtoil	.ets_vs1	.3604	228 .	100654	3.58	0.000	.162	9602	.5578853	
	advice2	.1907	452 .0	968147	1.97	0.049	.000	8146	.3806757	
landh	oldings	065	619 .0	351863	-1.86	0.062	134	5472	.0034093	
hhh	_mature	.0412	775.1	.060595	0.39	0.697	166	7895	.2493445	
hhh	_female	.010	044 .1	911317	0.05	0.958	364	9172	.3850051	
	_cons	-1.800	801 .1	246846	-14.44	0.000	-2.04	5406	-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.

regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female

. regress haz	ex_r sa	afewater	impro	/edtoilets	_vs1 a	advice2	2 landho	ldings	hhh_ma	ature hhh_f	emale
Source		SS	c	if M:	s	Numbe	er of ob	s =	1,3	305	
Model	120.	572361		7 17.224	6229	Prob	> F	=	0.0	.75 000	
Residual	3882	.54028	1,29	97 2.9934	7747	R-squ Adj R	lared (-square	= d =	0.03 0.03	301 249	
Total	4003	.11264	1,30	3.0698	7166	Root	MSE	=	1.7	302	
	haz	Coeffi	cient	Std. err.		t P	?> t	[95%	conf.	interval]	
	ex_r	.004	6134	.0012492	3	.69 0	.000	.002	1627	.0070641	
safe	ewater	.078	9408	.1091627	0	.72 0	.470	13	5214	.2930956	
improvedtoilet	s_vs1	.360	4228	.100654	3	.58 0	.000	.1629	9602	.5578853	
ac	lvice2	.190	7452	.0968147	1	.97 0	.049	.000	3146	.3806757	
landhol	ldings	06	5619	.0351863	-1	.86 0	.062	1340	5472	.0034093	
hhh_n	nature	.041	2775	.1060595	0	.39 0	.697	166	7895	.2493445	
hhh_ f	Female	.01	0044	.1911317	0	.05 0	.958	3649	9172	.3850051	
	_cons	-1.80	0801	.1246846	-14	.44 0	.000	-2.04	5406	-1.556195	

Г

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

regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female

. regress haz e	ex_r sa	fewater	improv	edtoilets_	_vs1 a	advice2	landhol	dings	hhh_ma	ature hhh_f	emale
Source		SS	ď	F MS	5	Numbe	r of obs	=	1,3	305	
Model	120.5	72361		7 17.2246	5229	Prob	1297) > F	=	0.00	.75 900	
Residual	3882.	54028	1,29	7 2.99347	7747	R-squ Adj R	ared -squared	=	0.03 0.03	301 249	
Total	4003.	11264	1,30	4 3.06987	7166	Root	MSE	=	1.73	302	
	haz	Coeffi	cient :	Std. err.		t P	> t	[95%	conf.	interval]	
safew	ex_r	.004	6134 9408	.0012492	3. 0.	.69 Ø	.000	.0021	627	.0070641	
improvedtoilets	_vs1	.360	4228	.100654	3.	.58 0	.000	.1629	602	.5578853	
adv	vice2	.190	7452	.0968147	1.	.97 0	.049	.0008	146	.3806757	
landhold	lings	06	5619	.0351863	-1.	.86 0	.062	1346	472	.0034093	
hhh_ma	ture	.041	2775	.1060595	0.	.39 0	.697	1667	895	.2493445	
hhh_fe	emale	.01	0044	.1911317	0.	.05 0	.958	3649	172	.3850051	
	cons	-1.80	0801	.1246846	-14	.44 0	.000	-2.045	406	-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).

regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female

. regress haz	ex_r sa	afewater i	improved	ltoilets_v	vs1 advio	ce2 landho	ldings	hhh_m	ature hhh_female
Source		SS	df	MS	Nur	mber of ob	s =	1,	305
Model	120.	572361	7	17.22462	— F() 29 Pro	7, 1297) ob > F	=	5 0.0	.75 000
Residual	3882	. 54028	1,297	2.993477	47 R-9	squared	=	0.0	301
Total	4003	.11264	1,304	3.069871	.66 Roo	ot MSE	=	1.7	302
						D	505%		
	haz	Coettic:	lent St	d. err.	t	P> t	[95%	cont.	intervalj
	ex_r	.00461	.134 .0	012492	3.69	0.000	.0021	627	.0070641
safe	ewater	.07894	.108	.091627	0.72	0.470	135	214	.2930956
improvedtoile	ts vs1	.36042	228 .	100654	3.58	0.000	.1629	602	.5578853
ad	dvice2	.19074	152 .0	968147	1.97	0.049	.0008	146	.3806757
landho.	ldings	0656	519 .0	351863	-1.86	0.062	1346	472	.0034093
hhh_r	mature	.04127	75.1	.060595	0.39	0.697	1667	895	.2493445
hhh_	female	.0100	944 .1	911317	0.05	0.958	3649	172	.3850051
	_cons	-1.8008	301 .1	246846	-14.44	0.000	-2.045	406	-1.556195

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

regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female

. regres	s haz	ex_r s	afewater	improve	edtoilets	_vs1	advice	2 landho	ldings	hhh_m	ature hhh <u></u>	_female
So	urce		SS	d	F M:	s	Numbe	er of ob	5 =	1,	305	
				-			F(7,	1297)	=	5	.75	
M	odel	120.	572361		7 17.224	6229	Prob	> F	=	0.0	000	
Resi	dual	3882	.54028	1,297	7 2.9934	7747	R-squ	uared	=	0.0	301	
							Adj H	R-square	= k	0.0	249	
Т	otal	4003	.11264	1,304	4 3.0698	7166	Root	MSE	=	1.7	302	
improved	safe	haz ex_r ewater	Coeffi .004 .078	cient 9 6134 9408 4228	Std. err. .0012492 .1091627 .100654	3	t 1 3.69 (3.72 (3.58 (>> t 0.000 0.470	[95% .002: 13! .1629	conf. 1627 5214 9602	interval .007064 .2930950	- - 1 3
1	20110	dvice2	190	7452	0968147	1	97 0	0.000	000	8146	380675	7
	au		.190	FC10	0351963	1		0.049	124	5140	. 380075	, ,
1	anuno.	Laings	00	2013	.0351805			0.002	1540	0472	.003409	-
	nnn_r	nature	.041	2775	.1060595	e	0.39 (0.697	166	/895	.249344	5
	hhh_ i	female	.01	0044	.1911317	e	9.05 (9.958	3649	9172	.385005	1
		_cons	-1.80	0801	.1246846	-14	1.44 (9.000	-2.04	5406	-1.55619	5

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.

regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female

. regress haz	ex_r sa	afewater	improved	toilets_\	/s1 advi	ce2 landho	ldings	hhh_m	ature hhh_fo	emale
Source		SS	df	MS	Nu	mber of ob	s =	1,	305	
Model	120.	572361	7	17.22462	— F(2 29 Pr	7, 1297) ob > F	=	5 0.0	.75 000	
Residual	3882	. 54028	1,297	2.993477	747 R-	squared	=	0.0	301	
Total	4003	.11264	1,304	3.069871	166 Ro	ot MSE	u = =	1.7	302	
	haz	Coeffic	ient St	d. err.	t	P> t	[95%	conf.	interval]	
	ex_r	.0046	134 .0	012492	3.69	0.000	.0021	627	.0070641	
safe	ewater ts vs1	.0789	408 .1	1091627	0.72 3.58	0.470 0.000	135	214 602	.2930956 .5578853	
a	dvice2	.1907	452 .6	968147	1.97	0.049	.0008	146	.3806757	
landho.	ldings	065	619 .0	351863	-1.86	0.062	1346	472	.0034093	
hhh_r	nature	.0412	.775	060595	0.39	0.697	1667	895	.2493445	
hhh_t	female	.010	044 .1	911317	0.05	0.958	3649	172	.3850051	
	_cons	-1.800	801 .1	246846	-14.44	0.000	-2.045	406	-1.556195	

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

regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female

 regress has 	z ex_r s	afewater :	improved	toilets_\	/s1 advi	ce2 landho	ldings	hhh_m	ature hhh_fema]	le
Source		SS	df	MS	Nur	mber of ob	s =	1,	305	
Model	120.	572361	7	17.22462	— F() 2 29 Pro	7, 1297) ob > F	=	5 0.0	.75 900	
Residual	3882	.54028	1,297	2.993477	747 R-:	squared	= - b	0.0	301	
Total	4003	.11264	1,304	3.069871	L66 Roo	ot MSE	=	1.7	302	
	haz	Coeffic	ient St	d. err.	t	P> t	[95%	conf.	interval]	
	ex_r	.0046	134 .0	012492	3.69	0.000	.0021	627	.0070641	
sa	fewater	.0789	408 .1	091627	0.72	0.470	135	214	.2930956	
improvedtoil	ets vsl	.3604	228	100654	3.58	0.000	.1629	602	.5578853	
	advice2	.1907	452 .0	968147	1.97	0.049	.0008	146	.3806757	
landh	oldings	065	619 .6	351863	-1.86	0.062	1346	472	.0034093	
hhh	_mature	.0412	775 .1	060595	0.39	0.697	1667	895	.2493445	
hhh	female	.010	044 .1	911317	0.05	0.958	3649	172	.3850051	
	_cons	-1.800	801 .1	246846	-14.44	0.000	-2.045	406	-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)

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

▶ Now, let's add province. What type of variable is province?

codebook prov - categorical/numeric/byte

- ▶ 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

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 province

. regress haz	ex_r sa	afewater impro	ovedtoilets	s_vs1 advi	ce2 landho	ldings hhh_m	ature hhh_female	2
Source		SS	df M	1S Nui	mber of ob	s = 1,	305	
Mada 1	224	125425	0 07 646	— F(3	8, 1296)	= 9	.47	
Model	221	125425 09733 1 ⁷	8 27.646	06/82 Pr	DD > F	= 0.0	000	
RESIDUAL	5761	.98/22 1,2	290 2.9182	-/ 10001	i R-squared	- 0.0	494	
Total	4003	.11264 1.3	304 3.0698	37166 Ro	ot MSE	= 1.7	083	
		· · · · · · · · · · · · · · · · · · ·						
	haz	Coefficient	Std. err.	t	P> t	[95% conf.	interval]	
	ex_r	.0041256	.0012362	3.34	0.001	.0017005	.0065508	
safe	ewater	0140756	.10894	-0.13	0.897	2277936	.1996425	
improvedtoilet	ts_vs1	.327707	.0995365	3.29	0.001	.1324366	.5229774	
ac	dvice2	.1516445	.0958214	1.58	0.114	0363376	.3396266	
landhol	ldings	0137426	.0358475	-0.38	0.702	084068	.0565829	
hhh_r	nature	.0103881	.1048496	0.10	0.921	1953055	.2160817	
hhh_t	female	.116171	.1895773	0.61	0.540	255741	.4880829	
pro	ovince	.063776	.0108647	5.87	0.000	.0424617	.0850903	
-	_cons	-2.290256	.1486872	-15.40	0.000	-2.58195	-1.998562	

regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female province

Source		SS	df	MS	5	Num	ber of obs	=	1,	305
			_			F(8	3, 1296)	=	9	.47
Model	221.	125425	8	27.6406	5/82	Pro)	=	0.0	000
Residual	3781	.98722 1,	296	2.91826	0001	R-s	quared	=	0.0	552
						Adj	R-squared	=	0.0	494
Total	4003	.11264 1,	304	3.06987	7166	Roc	ot MSE	=	1.7	083
	haz	Coefficient	St	d. err.		t	P> t	[95%	conf.	interval]
	ex_r	.0041256	.0	012362	3.	.34	0.001	.001	7005	.0065508
safe	ewater	0140756		.10894	-0.	13	0.897	227	7936	.1996425
mprovedtoilet	ts_vs1	.327707	.0	995365	3.	29	0.001	.132	4366	.5229774
ac	dvice2	.1516445	.0	958214	1.	58	0.114	036	3376	.3396266
landhol	ldings	0137426	.0	358475	-0.	38	0.702	08	4068	.0565829
hhh r	nature	.0103881	.1	048496	0.	10	0.921	195	3055	.2160817
hhh_t	Female	.116171	.1	895773	0.	61	0.540	25	5741	.4880829
pro	ovince	.063776	.0	108647	5.	87	0.000	.042	4617	.0850903
	cons	-2,290256	.1	486872	-15	40	0.000	-2.5	8195	-1,998562

How would we interpret the coefficient on "province"?

regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female province

> ce					
Source	SS	df MS	Number of obs	= 1,	305
			— F(8, 1296)	= 9	.47
Model 22:	.125425	8 27.640678	82 Prob > F	= 0.0	0000
Residual 378	31.98722 1,2	296 2.918200	01 R-squared	= 0.0	9552
			— Adj R-squared	= 0.0	9494
Total 400	3.11264 1,3	304 3.0698710	66 Root MSE	= 1.7	7083
ha:	Coefficient	Std. err.	t P> t	[95% conf.	interval]
ex_r	.0041256	.0012362	3.34 0.001	.0017005	.0065508
safewate	0140756	.10894	-0.13 0.897	2277936	.1996425
<pre>improvedtoilets_vs</pre>	. 327707	.0995365	3.29 0.001	.1324366	.5229774
advice	.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

. regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female provid

How would we interpret the coefficient on "province"?

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 provint

SourceSSdfMSNumber of obs=1,305Model221.125425827.6406782Prob > F=0.0000Residual3781.987221,2962.91820001R-squared=0.0552Total4003.112641,3043.06987166Root MSE=1.7083hazCoefficient Std. err.tP> t [95% conf. intervalex_r.0041256.00123623.340.001.0017005.00655safewater.0140756.10894-0.130.8972277936.19964advice2.1516445.09582141.580.1140363376.33962landholdings0137426.0358475-0.380.702084068.05588hhh_mature.0103881.10484960.100.9211953055.21608hhh female.116171.18957730.610.540255741.48808province.063776.01086475.870.000.0424617.08509	> ce										
Model221.125425827.6406782Prob > F= 0.0000 Residual3781.987221,2962.91820001R-squared= 0.0552 Total4003.112641,3043.06987166Root MSE=1.7083hazCoefficientStd. err.tP> t [95% conf. intervaex_r.0041256.00123623.340.001.0017005.00655safewater0140756.10894-0.130.8972277936.19964improvedtoilets_vs1.327707.09953653.290.001.1324366.52297advice2.1516445.09582141.580.1140363376.33962landholdings0137426.0358475-0.380.702084068.05658hhh_mature.0103881.10484960.100.9211953055.21608hhh female.116171.18957730.610.540255741.48808province.063776.01086475.870.000.0424617.08509	Source		SS	df	= MS	5	Numbe	er of obs	s =	1,	305
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 .00655 safewater 0140756 .10894 -0.13 0.897 2277936 .19964 improvedtoilets_vs1 .327707 .0995365 3.29 0.001 .1324366 .52297 advice2 .1516445 .0958214 1.58 0.114 0363376 .33962 landholdings 0137426 .0358475 -0.38 0.702 084068 .05658 hhh_mature .0103881 .1048496 0.10 0.921 1953055 .21608 hhh female .116171 .1895773 0.61 0.540 255741 .48808							F(8,	1296)	=	9	.47
Residual 3781.98722 1,296 2.91820001 R-squared = 0.0552 Total 4003.11264 1,304 3.06987166 Root MSE = 1.7083 haz Coefficient Std. err. t P> t [95% conf. interva ex_r .0041256 .0012362 3.34 0.001 .0017005 .00655 safewater 0140756 .10894 -0.13 0.897 2277936 .19964 improvedtoilets_vs1 .327707 .0995365 3.29 0.001 .1324366 .52297 advice2 .1516445 .0958214 1.58 0.114 0363376 .33962 landholdings 0137426 .0358475 -0.38 0.702 084068 .05658 hhh_mature .0103881 .1048496 0.10 0.921 1953055 .21608 hhh female .116171 .1895773 0.61 0.540 255741 .48808 province .063776 .0108647 5.87 0.000 .0424617 .08509	Model	221.1	L25425	8	27.6406	5782	Prob	> F	=	0.0	000
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 .00655 safewater 0140756 .10894 -0.13 0.897 2277936 .19964 improvedtoilets_vs1 .327707 .0995365 3.29 0.001 .1324366 .52297 advice2 .1516445 .0958214 1.58 0.114 0363376 .33962 landholdings 0137426 .0358475 -0.38 0.702 084068 .05658 hhh_mature .0103881 .1048496 0.10 0.921 1953055 .21608 hhh female .116171 .1895773 0.61 0.540 255741 .48808 province .063776 .0108647 5.87 0.000 .0424617 .08509	Residual	3781	98722	1,296	2.91826	001	R-squ	uared	=	0.0	552
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 .00655 safewater 0140756 .10894 -0.13 0.897 2277936 .19964 improvedtoilets_vs1 .327707 .0995365 3.29 0.001 .1324366 .52297 advice2 .1516445 .0958214 1.58 0.114 0363376 .33962 landholdings 0137426 .0358475 -0.38 0.702 084068 .05658 hhh_mature .0103881 .1048496 0.10 0.921 1953055 .21608 hhh female .116171 .1895773 0.61 0.540 255741 .48808 province .063776 .0108647 5.87 0.000 .0424617 .08509							Adj H	R-squared	d =	0.0	494
haz Coefficient Std. err. t P> t [95% conf. interval ex_r .0041256 .0012362 3.34 0.001 .0017005 .00655 safewater 0140756 .10894 -0.13 0.897 2277936 .19964 improvedtoilets_vs1 .327707 .0995365 3.29 0.001 .1324366 .52297 advice2 .1516445 .0958214 1.58 0.114 0363376 .33962 landholdings 0137426 .0358475 -0.38 0.702 084068 .05658 hhh_mature .0103881 .1048496 0.10 0.921 1953055 .21608 hhh female .116171 .1895773 0.61 0.540 255741 .48808 province .063776 .0108647 5.87 0.000 .0424617 .08509	Total	4003	11264	1,304	3.06987	166	Root	MSE	=	1.7	083
ex_r .0041256 .0012362 3.34 0.001 .0017005 .00655 safewater 0140756 .10894 -0.13 0.897 2277936 .19964 improvedtoilets_vs1 .327707 .0995365 3.29 0.001 .1324366 .52297 advice2 .1516445 .0958214 1.58 0.114 0363376 .33962 landholdings 0137426 .0358475 -0.38 0.702 084068 .05658 hhh_mature .0103881 .1048496 0.10 0.921 1953055 .21608 hhh female .116171 .1895773 0.61 0.540 255741 .48808 province .063776 .0108647 5.87 0.000 .0424617 .08509		haz	Coeffic	ient S	std. err.		t F	P> t	[95%	conf.	interval]
safewater 0140756 .10894 -0.13 0.897 2277936 .19964 improvedtoilets_vs1 .327707 .0995365 3.29 0.001 .1324366 .52297 advice2 .1516445 .0958214 1.58 0.114 0363376 .33962 landholdings 0137426 .0358475 -0.38 0.702 084068 .05658 hhh_mature .0103881 .1048496 0.10 0.921 1953055 .21608 hhh female .116171 .1895773 0.61 0.540 255741 .48808 province .063776 .0108647 5.87 0.000 .0424617 .08509		ex_r	.0041	256 .	0012362	з.	34 (9.001	.001	7005	.0065508
<pre>improvedtoilets_vs1 .327707 .0995365 3.29 0.001 .1324366 .52297 advice2 .1516445 .0958214 1.58 0.1140363376 .33962 landholdings0137426 .0358475 -0.38 0.702084068 .05658 hhh_mature .0103881 .1048496 0.10 0.9211953055 .21608 hhh female .116171 .1895773 0.61 0.540255741 .48808 province .063776 .0108647 5.87 0.000 .0424617 .08509</pre>	saf	ewater	0140	756	.10894	-0.	13 🤅	0.897	227	7936	.1996425
advice2 .1516445 .0958214 1.58 0.114 0363376 .33962 landholdings 0137426 .0358475 -0.38 0.702 084068 .05658 hhh_mature .0103881 .1048496 0.10 0.921 1953055 .21608 hhh female .116171 .1895773 0.61 0.540 255741 .48808 province .063776 .0108647 5.87 0.000 .0424617 .08509	improvedtoile [.]	ts_vs1	.327	707 .	0995365	3.	29 🤅	9.001	.1324	4366	.5229774
landholdings 0137426 .0358475 -0.38 0.702 084068 .05658 hhh_mature .0103881 .1048496 0.10 0.921 1953055 .21608 hhh female .116171 .1895773 0.61 0.540 255741 .48808 province .063776 .0108647 5.87 0.000 .0424617 .08509	a	dvice2	.1516	445.	0958214	1.	58 🤅	0.114	0363	3376	.3396266
hhh_mature .0103881 .1048496 0.10 0.921 1953055 .21608 hhh female .116171 .1895773 0.61 0.540 255741 .48808 province .063776 .0108647 5.87 0.000 .0424617 .08509	landho	ldings	0137	426.	0358475	-0.	38 🤅	9.702	084	4068	.0565829
hhh female .116171 .1895773 0.61 0.540255741 .48808 province .063776 .0108647 5.87 0.000 .0424617 .08509	hhh_i	mature	.0103	881 .	1048496	0.	10 0	9.921	1953	3055	.2160817
province .063776 .0108647 5.87 0.000 .0424617 .08509	h <u>hh</u>	female	.116	171 .	1895773	0.	61 🤅	0.540	25	5741	.4880829
	pro	ovince	.063	776.	0108647	5.	87 🤅	0.000	.0424	4617	.0850903
_cons -2.290256 .1486872 -15.40 0.000 -2.58195 -1.9985		_cons	-2.290	256 .	1486872	-15.	40 🤅	0.000	-2.58	8195	-1.998562

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

regress haz ex_r safewater improvedtoilets_vs1 advice2 landholdings hhh_mature hhh_female province

Source		SS	df	MS		Number of	obs	=	1,	305
Madal	221 1	125425	•	27 6406	707	F(8, 1296 Brob > F)	=	9	.47
Residual	3781	98722 1	0 296	27.0400	702 001	Prou > r R-squared		-	0.0	552
Residual	5701		,250	2.51020		Adi R-sau	ared	-	0.0	494
Total	4003	.11264 1	, 304	3.06987	166	Root MSE		=	1.7	083
	haz	Coefficien	t St	d. err.	t	P> t		[95%	conf.	interval]
	ex r	.0041256	.0	012362	3.3	4 0.001		.001	7005	.0065508
safe	ewater	0140756		.10894	-0.1	3 0.897	-	.227	7936	.1996425
provedtoile	ts_vs1	.327707	.0	995365	3.2	9 0.001		.1324	4366	.5229774
a	dvice2	.1516445	.0	958214	1.5	8 0.114	-	.0363	3376	.3396266
landho	ldings	0137426	.0	358475	-0.3	8 0.702		084	4068	.0565829
hhh_r	nature	.0103881	.1	.048496	0.1	0.921	-	.1953	3055	.2160817
hhh ·	Female	.116171	.1	.895773	0.6	1 0.540		25	5741	.4880829
pro	ovince	.063776	. e	108647	5.8	7 0.000		.0424	4617	.0850903
	_cons	-2.290256	.1	486872	-15.4	0.000		-2.58	8195	-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!

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

- 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

- 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)

- 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

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

 regress haz 	ex_r safewater	improved	toile	ts_vs1 a	dvice2 la	andholo	dings	hhh_mature h	hh_female i.	.provi
Source	SS	df		MS	Number o	of obs	=	1,305		
				4.2.6.0.5	F(20, 12	284)	=	7.05		
Model	396.2/3//	20	19.8	136885	Prob > F		=	0.0000		
Residual	3606.83887	1,284	2.86	906454	R-square	ea	=	0.0990		
Total	4003.11264	1,304	3.00	987166	Root MSE	luared	=	1.676		
	haz	Coeffic	ient	Std. er	r. t	: P:	> t	[95% conf	. interval]	
	ex_r	.0020	303	.0013157	7 1.5	54 0	.123	0005508	.0046114	
	safewater	.0192	869	.1093072	2 0.1	L8 Ø	.860	1951534	.2337272	
improv	vedtoilets_vs1	.2902	912	.1019296	5 2.8	85 0	.004	.0903243	.4902581	
	advice2	.0962	266	.106872	5 0.9	0 0	.368	1134372	.3058905	
	landholdings	.0405	974	.037919	5 1.6	97 0	.285	0337936	.1149885	
	hhh_mature	0140	334	.1038584	-0. 1	L4 Ø	.893	2177842	.1897175	
	hhh_female	.0317	556	.187157	5 0.1	L7 0	.865	3354125	.3989236	
	province									
Cent	tral Province	.5943	186	.2475298	3 2.4	0 0	.016	.1087114	1.079926	
Chimbu (Sir	mbu) Province	.4455	552	.2231952	2 2.6	0 0	.046	.007688	.8834225	
East New Brit	tain Province	.9316	815	.24660	5 3.7	78 0	.000	.4477806	1.415582	
East Se	epik Province	.3376	934	.2644979	9 1.2	28 0	.202	1812021	.8565889	
Eastern Highla	ands Province	.7891	199	.2534324	i 3.1	1 0	.002	.2919328	1.286307	
Jiv	waka Province	.5610	519	.2476679	2.2	27 0	.024	.0751739	1.04693	
Mad	dang Province	.7052	731	.2313612	2 3.6	95 0	.002	.2513857	1.15916	
Moi	robe Province	4243	679	.2527177	7 -1.6	58 Ø	.093	9201529	.0714171	
West	tern Province	1.267	225	.1785732	2 7.1	LØ Ø	.000	.9168979	1.617553	
Weste	ern Highlands	.5562	394	.2187539	2.5	64 0	.011	.1270851	.9853938	
(Gulf Province	1.293	717	.234365	5 5.5	52 0	.000	.833936	1.753499	
	Oro Province	1.167	969	.2215291	L 5.2	27 0	.000	.7324708	1.601668	
Milne	Bay Province	1.148	766	.2669933	4. 3	80 0	.000	.6249753	1.672557	

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

. regress haz	ex_r safewater	improved	toile	ets_vs1 a	dvice2	land	holdings	hhh_mature h	hh_female i.prov
Source	SS	df		MS	Number	of	obs =	1,305	
Model	206 27277	20	10 0	126995	F(20,	1284) =	7.05	
Posidual	390.2/3//	1 20	19.0	006464	Prod >	г nod	=	0.0000	
Residual	5000.85887	1,204	2.00	500454	Adi D_	caua	nod -	0.0950	
Total	4003.11264	1,304	3.06	987166	Root M	SE	=	1.676	
	haz	Coeffic	ient	Std. er	r.	t	P> t	[95% conf	. interval]
	ex_r	.0020	303	.001315	7 1	.54	0.123	0005508	.0046114
	safewater	.0192	869	.109307	2 0	.18	0.860	1951534	.2337272
impro	vedtoilets_vs1	.2902	912	.101929	6 2	.85	0.004	.0903243	.4902581
	advice2	.0962	266	.106872	5 0	.90	0.368	1134372	.3058905
	landholdings	.0405	974	.037919	5 1	.07	0.285	0337936	.1149885
	hhh_mature	0140	334	.103858	4 -0	.14	0.893	2177842	.1897175
	hhh_female	.0317	556	.187157	50	.17	0.865	3354125	.3989236
	province								
Con	tral Province	59/3	186	247529	8 2	10	0 016	108711/	1 079926
Chimbu (Si	mbu) Province		552	223195	2 2	. . 00	0.010	007688	8834225
Fast New Bri	tain Province	.9316	815	.2466	6 3	.78	0.000	. 4477806	1.415582
East S	enik Province	.3376	934	.264497	9 1	.28	0.202	1812021	.8565889
Eastern Highl	ands Province	.7891	199	.253432	4 3	.11	0.002	.2919328	1,286307
Ji	waka Province	.5610	519	.247667	9 2	.27	0.024	.0751739	1.04693
Ma	dang Province	.7052	731	.231361	2 3	.05	0.002	.2513857	1.15916
Mo	robe Province	4243	679	.252717	7 -1	.68	0.093	9201529	.0714171
Wes	tern Province	1.267	225	.178573	2 7	.10	0.000	.9168979	1.617553
West	ern Highlands	.5562	394	.218753	92	.54	0.011	.1270851	.9853938
	Gulf Province	1.293	717	.234365	5 5	.52	0.000	.833936	1.753499
	Oro Province	1.167	069	.221529	1 5	.27	0.000	.7324708	1.601668
Milne	Bay Province	1.148	766	.266993	3 4	.30	0.000	.6249753	1.672557
1		1							

Are any of the new province dummy variables significant?

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

 regress haz 	ex_r safewater	improved	toile	ets_vs1	advice2	landho	oldings	hhh_mature h	hh_female i	.provinc
Source	ss	df		MS	Number	of ol	bs =	1,305		
					F(20,	1284)	=	7.05		
Model	396.27377	20	19.8	3136885	Prob >	F	=	0.0000		
Residual	3606.83887	1,284	2.86	9906454	R-squa	red	=	0.0990		
					Adj R-	square	ed =	0.0850		
Total	4003.11264	1,304	3.00	5987166	Root M	SE	=	1.676		
	haz	Coeffic	ient	Std. e	rr.	t	P> t	[95% conf	. interval]	
	ex_r	.0020	303	.00131	57 1	.54	0.123	0005508	.0046114	-
	safewater	.0192	869	.10930	72 0	.18	0.860	1951534	.2337272	
improv	vedtoilets_vs1	.2902	912	.10192	96 2	.85	0.004	.0903243	.4902581	
	advice2	.0962	266	.10687	25 0	.90	0.368	1134372	.3058905	
	landholdings	.0405	974	.03791	95 1	.07	0.285	0337936	.1149885	
	hhh_mature	0140	334	.10385	84 -0	.14	0.893	2177842	.1897175	
	hhh_female	.0317	556	.18715	75 0	.17	0.865	3354125	.3989236	
	province									
Cent	tral Province	.5943	186	.24752	98 2	.40	0.016	.1087114	1.079926	i
Chimbu (Sir	mbu) Province	.4455	552	.22319	52 2	.00	0.046	.007688	.8834225	
East New Bri	tain Province	.9316	815	.246	f/6 3	.78	0.000	.4477806	1.415582	
East Se	epik Province	.3376	i934	.26449	7у 1	.28	0.202	1812021	.8565889)
Eastern Highla	ands Province	.7891	199	.25343	24 3	.11	0.002	.2919328	1.286307	,
Jiv	waka Province	.5610	519	.24766	79 2	.27	0.024	.0751739	1.04693	4
Mae	dang Province	.7052	731	.23136	12 3	.05	0.002	.2513857	1.15916	
Мог	robe Province	4243	679	.25271	77 -1	.68	0.093	9201529	.0714171	
West	tern Province	1.267	225	.1785	5 2 7	.10	0.000	.9168979	1.617553	
West	ern Highlands	. 5562	394	.21875	39 2	.54	0.011	.1270851	.9853938	
(Gulf Province	1.293	3717	.23436	55 5	.52	0.000	.833936	1.753499	
	Oro Province	1.167	069	.2215	S 1 5	.27	0.000	.7324708	1.601668	
Milne	Bay Province	1.148	3766	.26695	53 4	.30	0.000	.6249753	1.672557	·

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

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

	. regress haz	ex_r safewater	improved	toile	ts_vs1	ad	vice2 landh	oldings	hhh_mature h	hh_female i	i.province
	Source	SS	df		MS	I	Number of c	obs =	1,305		
						1	F(20, 1284)	=	7.05		
	Model	396.27377	20	19.8	136885		Prob > F	=	0.0000		
	Residual	3606.83887	1,284	2.80	906454	1	R-squared	=	0.0990		
							Adj R-squar	red =	0.0850		
	Total	4003.11264	1,304	3.06	987166	I	Root MSE	=	1.676		
		haz	Coeffic	ient	Std.	err	. t	P> t	[95% conf	. interval	-]
		ev n	0020	303	0013	157	1 5/	A 123	- 0005508	994611/	-
		safewater	0192	869	1093	372	0 18	0.125	- 1951534	233727	,
	improv	vedtoilets vs1	2902	912	1010	296	2 85	0.000	0903243	4902581	-
	Tubio	advice?	.2502	266	1069	725	A 9A	0.004	- 113/372	305890	
		landholdings	0405	974	0379	195	1 07	0.285	- 0337936	114988	5
		hhh mature	0140	334	.1038	584	-0.14	0.893	2177842	.189717	5
		hhh female	.0110	556	.1871	575	0.17	0.865	- 3354125	3989236	5
									1000011200		-
		province									
	Cent	tral Province	. 5943	186	.2475	298	2.40	0.016	.1087114	1,079926	5
	Chimbu (Sir	nbu) Province	.4455	552	.2231	952	2.00	0.046	.007688	.883422	5
1	East New Brit	tain Province	.9316	815	.24	66,6	3.78	0.000	.4477806	1,415582	2
ľ	East Se	epik Province	.3376	934	.2644	97y	1.28	0.202	1812021	.8565889	Ð
	Eastern Highla	ands Province	.7891	199	.2534	324	3.11	0.002	.2919328	1.286307	7
	Jiv	waka Province	.5610	519	.2476	679	2.27	0.024	.0751739	1.04693	3
	Mac	dang Province	.7052	731	.2313	612	3.05	0.002	.2513857	1.15916	5
	Mor	robe Province	4243	679	.2527	177	-1.68	0.093	9201529	.0714171	L
	West	tern Province	1.267	225	.1785	732	7.10	0.000	.9168979	1.617553	3
	Weste	ern Highlands	.5562	394	.2187	539	2.54	0.011	.1270851	.9853938	3
	(Gulf Province	1.293	717	.2343	655	5.52	0.000	.833936	1.753499	Ð
		Oro Province	1.167	069	.2215	291	5.27	0.000	.7324708	1.601668	3
	Milne	Bay Province	1.148	766	.2669	933	4.30	0.000	.6249753	1.672557	7

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.

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

. regress haz	ex_r safewater	improved	ltoile	ts_vs1 a	dvice2 land	iholdings	hhh_mature h	hh_female i.provin
Source	SS	df		MS	Number of	obs =	1,305	
					F(20, 1284	+) =	7.05	
Model	396.27377	20	19.8	136885	Prob > F	=	0.0000	
Residual	3606.83887	1,284	2.80	906454	R-squared	=	0.0990	
/					Adj R-squa	ared =	0.0850	
Total	4003.11264	1,304	3.06	987166	Root MSE	=	1.676	
	haz	Coeffic	cient	Std. er	r. t	P> t	[95% conf	. interval]
	ex_r	.0026	3 303	.001315	7 1.54	0.123	0005508	.0046114
I	safewater	.0197	2869	.109307	2 0.18	0.860	1951534	.2337272
improv	vedtoilets_vs1	. 2907	2912	.101929	6 2.85	0.004	.0903243	.4902581
-	advice2	.0967	2266	.106872	.5 0.90	0.368	1134372	.3058905
	landholdings	.0405	5974	.037919	5 1.07	0.285	0337936	.1149885
I	hhh_mature	0146	ð334	.103858	4 -0.14	0.893	2177842	.1897175
l	hhh_female	.0317	/556	.187157	5 0.17	0.865	3354125	.3989236
1	province							
Cent	tral Province	.5943	3186	.247529	8 2.40	0.016	.1087114	1.079926
Chimbu (Sir	mbu) Province	.4455	5552	.223195	2 2.00	0.046	.007688	.8834225
East New Brit	tain Province	.9316	5815	.2466	6 3.78	0.000	.4477806	1.415582
East Sc	epik Province	.337€	á934	.264497	9 1.28	0.202	1812021	.8565889
Eastern Highl	ands Province	.7891	1199	.253432	.4 3.11	0.002	.2919328	1.286307
Ji	waka Province	.5616	ð519	.247667	9 2.27	0.024	.0751739	1.04693
Mar	dang Province	.7052	2731	.231361	.2 3.05	0.002	.2513857	1.15916
Mor	robe Province	4243	3679	.252717	7 -1.68	0.093	9201529	.0714171
West	tern Province	1.267	7225	.17857?	2 7.10	0.000	.9168979	1.617553
West	ern Highlands	. 5567	2394	. 21875?	м <u>2.54</u>	0.011	.1270851	.9853938
(Gulf Province	1.293	3717	.234365	5 5.52	0.000	.833936	1.753499
1	Oro Province	1.167	7069	.22152	1 5.27	0.000	.7324708	1.601668
Milne	Bay Province	1.145	8766	266997	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.

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

► 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	improved	toile	ets_vs1	adv	vice2 landh	oldings	hhh_mature h	hh_female	i.province
	Source	SS	df		MS	N	lumber of o	bs =	1,305		
						F	(20, 1284)	=	7.05		
	Model	396.27377	20	19.8	3136885	P	rob > F	=	0.0000		
	Residual	3606.83887	1,284	2.86	9906454	R	-squared	=	0.0990		
						А	dj R-squar	ed =	0.0850		
	Total	4003.11264	1,304	3.06	5987166	R	loot MSE	=	1.676		
			1								_
		haz	Coeffic	ient	Std.	err.	t	P> t	[95% conf	. interval]
		ex_r	.0020	303	.0013	157	1.54	0.123	0005508	.004611	4
		safewater	.0192	869	.1093	072	0.18	0.860	1951534	.233727	2
	improv	.2902	912	.1019	296	2.85	0.004	.0903243	.490258	1	
		advice2	.0962	266	.1068	725	0.90	0.368	1134372	.305890	5
		landholdings	.0405	974	.0379	195	1.07	0.285	0337936	.114988	5
		hhh_mature	0140	334	.1038	584	-0.14	0.893	2177842	.189717	5
		hhh_female	.0317	556	.1871	575	0.17	0.865	3354125	.398923	6
		province									
	Cent	tral Province	.5943	186	.2475	298	2.40	0.016	.1087114	1.07992	6
Chimbu (Simbu) Province		.4455	552	.2231	952	2.00	0.046	.007688	.883422	5	
	East New Bri	tain Province	.9316	815	.24	666	3.78	0.000	.4477806	1.41558	2
	East Se	epik Province	.3376	934	.2644	979	1.28	0.202	1812021	.856588	9
	Eastern Highlands Province		.7891	199	.2534	324	3.11	0.002	.2919328	1.28630	7
	Jiv	waka Province	.5610	519	.2476	679	2.27	0.024	.0751739	1.0469	3
	Mad	dang Province	.7052	731	.2313	612	3.05	0.002	.2513857	1.1591	6
	Mor	robe Province	4243	679	.2527	177	-1.68	0.093	9201529	.071417	1
	West	tern Province	1.267	225	.1785	732	7.10	0.000	.9168979	1.61755	3
	Weste	ern Highlands	.5562	394	.2187	539	2.54	0.011	.1270851	.985393	8
	(Gulf Province	1.293	717	.2343	655	5.52	0.000	.833936	1.75349	9
		Oro Province	1.167	069	.2215	291	5.27	0.000	.7324708	1.60166	8
	Milne	Bay Province	1.148	766	.2669	933	4.30	0.000	.6249753	1.67255	7
			1								

4. Ordinary least-squares (OLS) linear regression

 regress haz 	ex_r safewater	improved	toile	ets_vs1	adv	/ice2 landh	oldings	hhh_mature h	hh_female	i.province
Source	SS	df		MS	Ν	Number of c	obs =	1,305		
					F	(20, 1284)	=	7.05		
Model	396.27377	20	19.8	8136885	F	Prob > F	=	0.0000		
Residual	3606.83887	1,284	2.86	906454	F	R-squared	. =	0.0990		
					- /	Adj R-squar	red =	0.0850		
Total	4003.11264	1,304	3.06	987166	F	Root MSE	=	1.676		
	haz	Coeffic	ient	Std.	err.	. t	P> t	[95% conf	. interval	.]
	ex r	.0020	303	.0013	157	1.54	0.123	0005508	.004611	.4
	safewater	.0192	869	.1093	072	0.18	0.860	1951534	.233727	2
impro	vedtoilets_vs1	.2902	912	.1019	296	2.85	0.004	.0903243	.490258	1
	advice2	.0962	266	.1068	725	0.90	0.368	1134372	.305890	15
	landholdings	.0405	974	.0379	195	1.07	0.285	0337936	.114988	5
	hhh_mature	0140	334	.1038	584	-0.14	0.893	2177842	.189717	'5
	hhh_female	.0317	556	.1871	575	0.17	0.865	3354125	.398923	6
	province									
Cen	tral Province	.5943	186	.2475	298	2.40	0.016	.1087114	1.07992	6
Chimbu (Simbu) Province		.4455	552	.2231	952	2.00	0.046	.007688	.883422	.5
East New Bri	tain Province	.9316	815	.24	666	3.78	0.000	.4477806	1.41558	32
East S	epik Province	.3376	934	.2644	979	1.28	0.202	1812021	.856588	19
Eastern Highla	.7891	199	.2534	324	3.11	0.002	.2919328	1.28630)7	
Jiv	waka Province	.5610	519	.2476	679	2.27	0.024	.0751739	1.0469	3
Ma	dang Province	.7052	731	.2313	612	3.05	0.002	.2513857	1.1591	.6
Мо	robe Province	4243	679	.2527	177	-1.68	0.093	9201529	.071417	'1
Wes	tern Province	1.267	225	.1785	732	7.10	0.000	.9168979	1.61755	3
West	ern Highlands	.5562	394	.2187	539	2.54	0.011	.1270851	.985393	8
	Gulf Province	1.293	717	.2343	655	5.52	0.000	.833936	1.75349	19
	Oro Province	1.167	069	.2215	291	5.27	0.000	.7324708	1.60166	8
Milne	Bay Province	1.148	766	.2669	933	4.30	0.000	.6249753	1.67255	7

What's significant?

4. Ordinary least-squares (OLS) linear regression

regress haz	ex_r safewater	improved	toil	ets_vs1 a	advice2 land	dholdings	hhh_mature hh	h_female i.prov
Source	SS	df		MS	Number of	obs =	1,305	
					F(20, 1284	4) =	7.05	
Model	396.27377	20	19.8	8136885	Prob > F	=	0.0000	
Residual	3606.83887	1,284	2.8	9906454	R-squared	=	0.0990	
					Adj R-squa	ared =	0.0850	
Total	4003.11264	1,304	3.0	5987166	Root MSE	=	1.676	
	haz	Coeffic	ient	Std. e	rr. t	P> t	[95% conf.	interval]
	ex_r	.0020	303	.00131	57 1.54	0.123	0005508	.0046114
	safewater	.0192	869	.10930	0.18	0.860	1951534	.2337272
impro	vedtoilets_vs1	.2902	912	.101929	2.85	0.004	.0903243	.4902581
	advice2	.0962	266	.106872	25 0.90	0.368	1134372	.3058905
	landholdings	.0405	974	.037919	95 1.07	0.285	0337936	.1149885
	hhh_mature	0140	334	.10385	-0.14	0.893	2177842	.1897175
	hhh_female	.0317	556	.18715	0.17	0.865	3354125	.3989236
	province							
Cen	tral Province	.5943	186	.247529	2.40	0.016	.1087114	1.079926
Chimbu (Si	mbu) Province	.4455	552	.22319	2.00	0.046	.007688	.8834225
East New Bri	tain Province	.9316	815	.246	56 3.78	0.000	.4477806	1.415582
East S	epik Province	.3376	934	.26449	9 1.28	0.202	1812021	.8565889
astern Highl	ands Province	.7891	199	.253432	24 3.11	0.002	.2919328	1.286307
Ji	waka Province	.5610	519	.24766	79 2.27	0.024	.0751739	1.04693
Ma	dang Province	.7052	731	.23136	L2 3.05	0.002	.2513857	1.15916
Mo	robe Province	- 4743	679	25271	-1.68	0,093	- 9201529	.0714171
Wes	tern Province	1.267	225	.178573	32 7.10	0.000	.9168979	1.617553
West	ern Highlands	. 5562	394	.21875	39 2.54	0.011	.1270851	.9853938
(Gulf Province	1.293	717	.23436	55 5.52	0.000	.833936	1.753499
	Oro Province	1.167	069	.221529	5.27	0.000	.7324708	1.601668
Milne	Bay Province	1.148	766	.266993	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.