Empirical Result Presentation

Academic Writing

Wang, Yu

dr.yuwang@outlook.com

December 24, 2024

Overview

1. How to Present Empirical Models
   - Which Information Should Be Reported
   - Which Models Should Be Reported
2. How to Interpret Empirical Results
   - Interpreting Results Substantively
   - Common Mistakes and Awkward Expressions
3. How to Design Tables
   - Components of a Table
   - Design Tables Professionally
4. Tools for Automated Table Creation

Table of Contents

Sample Table (from *AER*)

TABLE 9—CROP YIELD, GROWTH CYCLE, AND LONG-TERM ORIENTATION: REGIONAL-LEVEL ANALYSIS BASED ON WVS

	Share of individuals in WVS region with long-term orientation											
	Whole world										Old World	
	Unweighted				Weighted: area				Weighted: area share		Area	Share
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Crop yield	0.049	0.046	0.053		0.097		0.032		0.031		0.039	0.032
	(0.012)	(0.013)	(0.017)		(0.033)		(0.012)		(0.013)		(0.015)	(0.013)
Crop growth cycle			−0.010		−0.047		−0.024		−0.036		−0.027	−0.036
			(0.012)		(0.021)		(0.010)		(0.009)		(0.009)	(0.008)
Crop yield (ancestors)				0.077		0.133		0.043		0.041		
				(0.020)		(0.032)		(0.017)		(0.017)		
Crop growth cycle (ancestors)				−0.012		−0.050		−0.027		−0.037		
				(0.013)		(0.018)		(0.009)		(0.009)		
Continental FE	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	No
Country FE	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Geographical controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Old World sample	No	No	No	No	No	No	No	No	No	No	Yes	Yes
Weighted by region area	No	No	No	No	Yes	Yes	Yes	Yes	No	No	Yes	No
Weighted by region's share of area	No	No	No	No	No	No	No	No	Yes	Yes	No	Yes
Adjusted R^2	0.22	0.25	0.25	0.28	0.28	0.37	0.72	0.72	0.86	0.86	0.72	0.86
Observations	1,356	1,356	1,356	1,356	1,356	1,356	1,356	1,356	1,356	1,356	1,143	1,143

Notes: The table shows the effect of a region's pre-1500CE potential crop yield and its post-1500CE change in the course of the Columbian Exchange (measured in calories per hectare per year) on the share of its population with long-term orientation, accounting of country fixed effects. Geographical controls are absolute latitude, mean elevation above sea level, terrain roughness, percentage of land within 100km of sea, landlocked dummy, and area suitable for agriculture. Columns 1–4 show the unweighted results; columns 5–8 weight observations according to the region's area; columns 9–10 weight observations according to the region's area as a share of the country's area; and columns 11–12 conduct the analysis for the Old World sample. All independent variables have been normalized by subtracting their mean and dividing by their standard deviation. Thus, all coefficients can be compared and show the effect of a one standard deviation increase in the independent variable on long-term orientation. Heteroskedasticity-robust standard error estimates clustered at the country level are reported in parentheses.

Parameter Estimates of Regressors (with Standard Errors)

1) Report parameter estimates and their standard errors.
   - Should individual intercepts be reported (in a fixed effect model)?

Parameter Estimates of Regressors (with Standard Errors)

1) Report parameter estimates and their standard errors.
   - Should individual intercepts be reported (in a fixed effect model)?
   - Do not report if parameter estimates of other regressors are of primary interest;
   - Report them if time-invariant unobserved heterogeneity (individual effects) is of main research interest.

Goodness-of-fit Measures

2) Report goodness-of-fit measures that examine the extant that the model fits data:

- F-test (or likelihood ratio test) to test the model and its significance (p-value);
- R^2 or adjusted R^2 in OLS and fixed effect models; theta θ and variance components $\hat{\sigma}_u$ estimated in a random effect model;
- Sum of squared errors (residual);
- Degrees of freedom (d.f.) for errors;
- Number of observations $N(= nT)$ and number of cross-sections n, etc.

In case of poor goodness-of-fit, you need to try other models.

Test Results

Report and interpret test results if available, e.g., in panel data modeling, report if fixed and/or random effect exists:

- F-test result for a fixed effect model;
- Breusch-Pagan LM test result for a random effect model;
- Hausman test result to determine which model shall be used when both fixed and random effects are statistically significant;
- Chow test result to show the poolability of data if you doubt constant slopes across individuals, etc.

Exercise 1

Find the 1) parameter estimates (and their s.e.); 2) goodness-of-fit measures; and 3) test results from the table below.

Table 5.1 Comparison of OLS, LSDV, and Within Effect Models

	OLS	LSDV	"Within"	.xtreg	.areg
Ouput index	.8827**	.9193**	.9193**	.9193**	.9193**
	(.0133)	(.0299)	(.0288)	(.0299)	(.0299)
Fuel price	.4540**	.4175**	.4175**	.4175**	.4175**
	(.0203)	(.0152)	(.0147)	(.0152)	(.0152)
Loading factor	-1.6275**	-1.0704**	-1.0704**	-1.0704**	-1.0704**
	(.3453)	(.2017)	(.1946)	(.2017)	(.2017)
Intercept (baseline)	9.5169**	9.7930**		9.7135**	9.7135**
	(.2292)	(.2637)		(.2296)	(.2296)
Airline 1 (dummy)		-.0871			
		(.0842)			
Airline 2 (dummy)		-.1283			
		(.0757)			
Airline 3 (dummy)		-.2960**			
		(.0500)			
Airline 4 (dummy)		.0975**			
		(.0330)			
Airline 5 (dummy)		-.0630**			
		(.0239)			
F-test (model)	2419.34**	3935.79**	3871.82**	3604.80**	3604.80**
Degrees of freedom	86	81	81	81	81
SSM (model)	112.7054	113.7483	39.0684	*39.0684*	*113.7483*
SSE (error/residual)	1.3354	.2926	.2926	*.2926*	*.2926*
Root MSE (SEE)	.1246	.0601	.0580	.0601	.0601
R^2	.9883	.9974	.9926	*.9926*	.9974
Adjusted R^2	.9879	.9972	.9923	*.9918*	.9972
F-test (fixed effect)				57.73**	57.732**
N	90	90	90	90	90

Source: http://pages.stern.nyu.edu/~wgreene/Text/tables/tablelist5.htm

* Standard errors in parenthesis; Statistics hidden in macros are italicized; Statistical significance: * <.05, ** <.01

Trade-off Between Robustness and Accuracy

- Determine and report an **"accurate"** model (the baseline model).

  If one model is "right," the other models are "wrong."
- Present model variations to check **robustness** of the baseline model results.

  Variations in estimators, control variable sets, variable definitions, subsamples, etc.

Table of Contents

Interpreting Empirical Results

Three questions to answer:

- What are the answers to your research questions?
- What is the relationship between your DV and the IV(s) you have chosen to examine?
- Does the model "fit" the observed data?

There are three tasks to fulfill ...

Interpreting Empirical Results

Task 1. Introduce the table explicitly.

- Point out to your readers that the table exists and indicate, briefly, its general content.
- e.g., "Table 1 shows the incomes earned by full-time workers in the United States."
- e.g., "In Table 1, I present the results of the three regressions that explore the relationship between income and education."

Interpreting Empirical Results

Task 2. Identify the main points made by the data in the table, the points that most closely correspond to your research question.

- Explicitly tell your readers the important realities that the data show.
- e.g, "Table 1 reveals several significant characteristics of our sample that could affect our results: one-third of women in the sample had less than a high-school education; nearly two-thirds were unmarried; and exactly one-half had at least one child under 3."
- e.g, "As expected, the coefficient on education is, in every regression, significant and positive."
- You may also wish to point out any counter-intuitive results or results that are especially large or small.

Interpreting Empirical Results

Task 3. Draw your reader's attention to the applicable numerical figures in the table.

- e.g, "As seen in column 1, the coefficient on education is 0.583 and is statistically significant at the 5 percent level."
- e.g., "For one unit increase in an IV, DV is expected to increase by X units(, holding all other variables constant)."(the *ceteris paribus* assumption is usually omitted).
- Do not simply report signs and magnitudes of coefficients, e.g., an independent variable is "significant," "negatively (or positively) related to ...", or "insignificantly related ...".
- Provide statistical significance in the table and the p-value in parenthesis at the end of the interpretation sentence.

Exercise 2

Table 1 OLS Estimates of the Effect of Education on Wages.
Dependent Variable: Log of Yearly Earnings, 1985–1995

	1	2	3	4
Years of Education	.091 (.001)	.031 (.003)	.086 (.002)	.027 (.005)
Ability Dummy		.251 (.010)		.301 (.010)
State Dummies Included?	No	No	Yes	Yes
No. of Observations	35,001	35,001	19,505	18,505
No. of Persons	5,505	5,505	4,590	4,590
Adjusted R^2	.50	.55	.76	.79

Note: Standard errors are in parentheses. Data are from the Tennessee Second Grade Ability Survey and Wage Follow-up, and include individuals evaluated between 1962 and 1971. The "ability dummy" equals 1 if the individual's second-grade teacher classified the individual as "able," and 0 otherwise. Each regression also includes yearly dummies, ten one-digit industry and twenty Census-defined occupation dummies, labor market experience (defined as one's age minus 6), experience squared, seniority on the current job, seniority squared, Census region of current residence, marital status, race, gender, and a dummy variable denoting whether the individual lives in a city of more than 100,000 persons. Columns 3 and 4 have fewer observations because the state of residence is not available for some individuals.

Interpret the regression results in the table. You should fulfill the "three tasks" in your interpretation.

Exercise 2

Table 1 OLS Estimates of the Effect of Education on Wages.
Dependent Variable: Log of Yearly Earnings, 1985–1995

	1	2	3	4
Years of Education	.091 (.001)	.031 (.003)	.086 (.002)	.027 (.005)
Ability Dummy		.251 (.010)		.301 (.010)
State Dummies Included?	No	No	Yes	Yes
No. of Observations	35,001	35,001	19,505	18,505
No. of Persons	5,505	5,505	4,590	4,590
Adjusted R^2	.50	.55	.76	.79

Note: Standard errors are in parentheses. Data are from the Tennessee Second Grade Ability Survey and Wage Follow-up, and include individuals evaluated between 1962 and 1971. The "ability dummy" equals 1 if the individual's second-grade teacher classified the individual as "able," and 0 otherwise. Each regression also includes yearly dummies, ten one-digit industry and twenty Census-defined occupation dummies, labor market experience (defined as one's age minus 6), experience squared, seniority on the current job, seniority squared, Census region of current residence, marital status, race, gender, and a dummy variable denoting whether the individual lives in a city of more than 100,000 persons. Columns 3 and 4 have fewer observations because the state of residence is not available for some individuals.

Sentence 1 (Task 1):
"Table 1 presents the OLS estimates of the effect of education on wages."

Exercise 2

Table 1 OLS Estimates of the Effect of Education on Wages.
Dependent Variable: Log of Yearly Earnings, 1985–1995

	1	2	3	4
Years of Education	.091 (.001)	.031 (.003)	.086 (.002)	.027 (.005)
Ability Dummy		.251 (.010)		.301 (.010)
State Dummies Included?	No	No	Yes	Yes
No. of Observations	35,001	35,001	19,505	18,505
No. of Persons	5,505	5,505	4,590	4,590
Adjusted R^2	.50	.55	.76	.79

Note: Standard errors are in parentheses. Data are from the Tennessee Second Grade Ability Survey and Wage Follow-up, and include individuals evaluated between 1962 and 1971. The "ability dummy" equals 1 if the individual's second-grade teacher classified the individual as "able," and 0 otherwise. Each regression also includes yearly dummies, ten one-digit industry and twenty Census-defined occupation dummies, labor market experience (defined as one's age minus 6), experience squared, seniority on the current job, seniority squared, Census region of current residence, marital status, race, gender, and a dummy variable denoting whether the individual lives in a city of more than 100,000 persons. Columns 3 and 4 have fewer observations because the state of residence is not available for some individuals.

Sentence 2 (Task 2):
"It shows that including a measure of ability in the wage equation dramatically lowers the predicted effect of education on earnings."

Exercise 2

Table 1 OLS Estimates of the Effect of Education on Wages. Dependent Variable: Log of Yearly Earnings, 1985–1995

	1	2	3	4
Years of Education	.091 (.001)	.031 (.003)	.086 (.002)	.027 (.005)
Ability Dummy		.251 (.010)		.301 (.010)
State Dummies Included?	No	No	Yes	Yes
No. of Observations	35,001	35,001	19,505	18,505
No. of Persons	5,505	5,505	4,590	4,590
Adjusted R^2	.50	.55	.76	.79

Note: Standard errors are in parentheses. Data are from the Tennessee Second Grade Ability Survey and Wage Follow-up, and include individuals evaluated between 1962 and 1971. The "ability dummy" equals 1 if the individual's second-grade teacher classified the individual as "able," and 0 otherwise. Each regression also includes yearly dummies, ten one-digit industry and twenty Census-defined occupation dummies, labor market experience (defined as one's age minus 6), experience squared, seniority on the current job, seniority squared, Census region of current residence, marital status, race, gender, and a dummy variable denoting whether the individual lives in a city of more than 100,000 persons. Columns 3 and 4 have fewer observations because the state of residence is not available for some individuals.

Sentence 3 (Task 3):
"Column 1 does not include an ability measure and indicates that a year of education raises wages by 9.1 percent, and this effect is statistically significant at the 1 percent level. Column 2 adds the ability measure; the education effect now drops to 3.1 percent. Columns 3 and 4 show that this general pattern is repeated even when state-level dummy variables are included."

Exercise 2

Table 1 OLS Estimates of the Effect of Education on Wages.
Dependent Variable: Log of Yearly Earnings, 1985–1995

	1	2	3	4
Years of Education	.091 (.001)	.031 (.003)	.086 (.002)	.027 (.005)
Ability Dummy		.251 (.010)		.301 (.010)
State Dummies Included?	No	No	Yes	Yes
No. of Observations	35,001	35,001	19,505	18,505
No. of Persons	5,505	5,505	4,590	4,590
Adjusted R^2	.50	.55	.76	.79

Note: Standard errors are in parentheses. Data are from the Tennessee Second Grade Ability Survey and Wage Follow-up, and include individuals evaluated between 1962 and 1971. The "ability dummy" equals 1 if the individual's second-grade teacher classified the individual as "able," and 0 otherwise. Each regression also includes yearly dummies, ten one-digit industry and twenty Census-defined occupation dummies, labor market experience (defined as one's age minus 6), experience squared, seniority on the current job, seniority squared, Census region of current residence, marital status, race, gender, and a dummy variable denoting whether the individual lives in a city of more than 100,000 persons. Columns 3 and 4 have fewer observations because the state of residence is not available for some individuals.

Sentence 4 (Task 2):
"The estimates in Table 1 are therefore consistent with the hypothesis that the OLS estimates suffer from an upward ability bias."

Statistical Significance

What's wrong with the following sentence?

- "This effect is statistically significant at 5% significant level."

Statistical Significance

- Do not say "significant level," "at 5% level," or "at the level of significance $\alpha = 5\%$," etc.;
- Say "significance level," "at the 5 percent level," or "at the 5 percent significance level;"
- Use a specific significance level (e.g., "at the 10 percent level") rather than "at the conventional level."

Hypothesis

What's wrong with the following sentence?

- "Our null hypothesis can be expressed as $b_1 = 0$."

Hypothesis

- Do not use "$b_1 = 0$" as the hypothesis;
- Use "$\beta_1 = 0$" as the hypothesis;
- A hypothesis is a conjecture about the unknown (e.g, α, β).
- We do not need to test $b_1 = 0$ as the b_1 is already known (estimated from the sample).

Parameter Estimates

What's wrong with the following sentence?

- "The coefficient of β_1 is 0.22."

Parameter Estimates

- Do not say "the coefficient of β_1;"
- Say "parameter estimates of β_1" or "the coefficient of IV1;"
- Do not say "beta" or "beta coefficient;"
- Say "standardized coefficient of IV."

p-values

What's wrong with the following sentence?

- "The p-value is significant at the 10 percent level."

p-values

- Do not say "the p-value is significant;"
- Say "the p-value is small enough to reject H_0" or "a small p-value suggests rejection of H_0;"
- A p-value itself is neither significant nor insignificant.

Reject or Do Not Reject the Null Hypothesis

What's wrong with the following sentence?

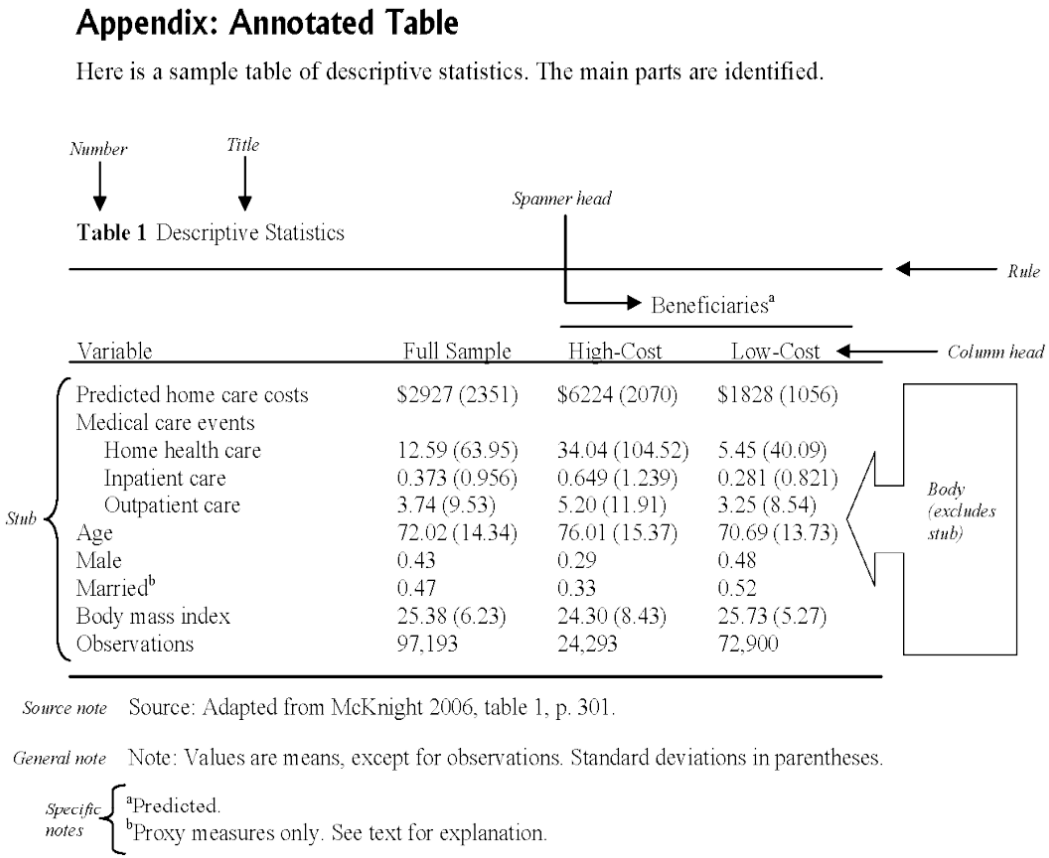
- "Therefore, we accept the null hypothesis."

Reject or Do Not Reject the Null Hypothesis

- Do not say "accept (or confirm)" the null hypothesis;
- Say "reject" or "do not reject" the null hypothesis;
- Do not say "I do not believe that the H_0 is true" or "the test provides decisive evidence that the H_0 is wrong;"
- Say "reject the H_0 at the 1 percent level;"
- No one knows if a H_0 is really true or wrong.

Table of Contents

Components of a Table



Appendix: Annotated Table

Here is a sample table of descriptive statistics. The main parts are identified.

Number *Title*

Table 1 Descriptive Statistics

Spanner head

Rule

Beneficiaries[a]

Variable	Full Sample	High-Cost	Low-Cost
Predicted home care costs	$2927 (2351)	$6224 (2070)	$1828 (1056)
Medical care events			
Home health care	12.59 (63.95)	34.04 (104.52)	5.45 (40.09)
Inpatient care	0.373 (0.956)	0.649 (1.239)	0.281 (0.821)
Outpatient care	3.74 (9.53)	5.20 (11.91)	3.25 (8.54)
Age	72.02 (14.34)	76.01 (15.37)	70.69 (13.73)
Male	0.43	0.29	0.48
Married[b]	0.47	0.33	0.52
Body mass index	25.38 (6.23)	24.30 (8.43)	25.73 (5.27)
Observations	97,193	24,293	72,900

Column head

Stub

Body (excludes stub)

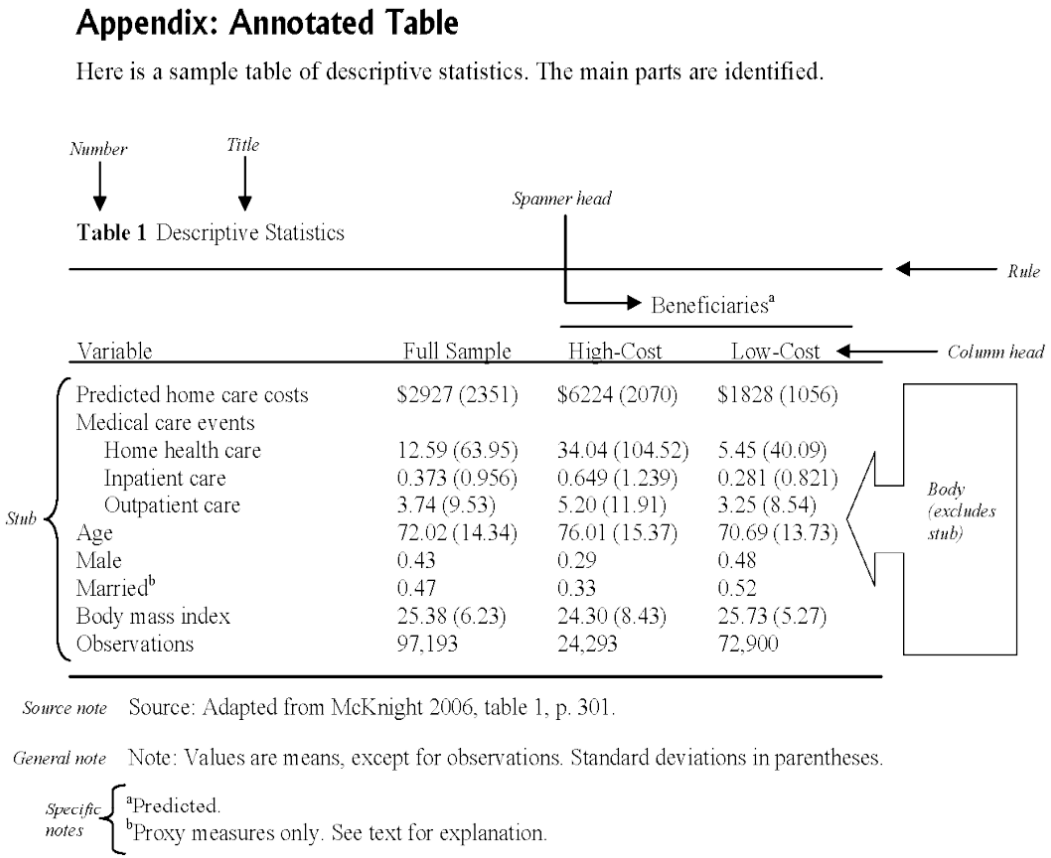
Source note Source: Adapted from McKnight 2006, table 1, p. 301.

General note Note: Values are means, except for observations. Standard deviations in parentheses.

Specific notes [a]Predicted.
[b]Proxy measures only. See text for explanation.

- Table number
- Title
- Column heads
- Stub
- Body
- Footnotes
- Rules

Components of a Table



Appendix: Annotated Table

Here is a sample table of descriptive statistics. The main parts are identified.

Number *Title*

Table 1 Descriptive Statistics

Spanner head

Rule

Beneficiaries[a]

Variable	Full Sample	High-Cost	Low-Cost
Predicted home care costs	$2927 (2351)	$6224 (2070)	$1828 (1056)
Medical care events			
Home health care	12.59 (63.95)	34.04 (104.52)	5.45 (40.09)
Inpatient care	0.373 (0.956)	0.649 (1.239)	0.281 (0.821)
Outpatient care	3.74 (9.53)	5.20 (11.91)	3.25 (8.54)
Age	72.02 (14.34)	76.01 (15.37)	70.69 (13.73)
Male	0.43	0.29	0.48
Married[b]	0.47	0.33	0.52
Body mass index	25.38 (6.23)	24.30 (8.43)	25.73 (5.27)
Observations	97,193	24,293	72,900

Column head

Stub

Body (excludes stub)

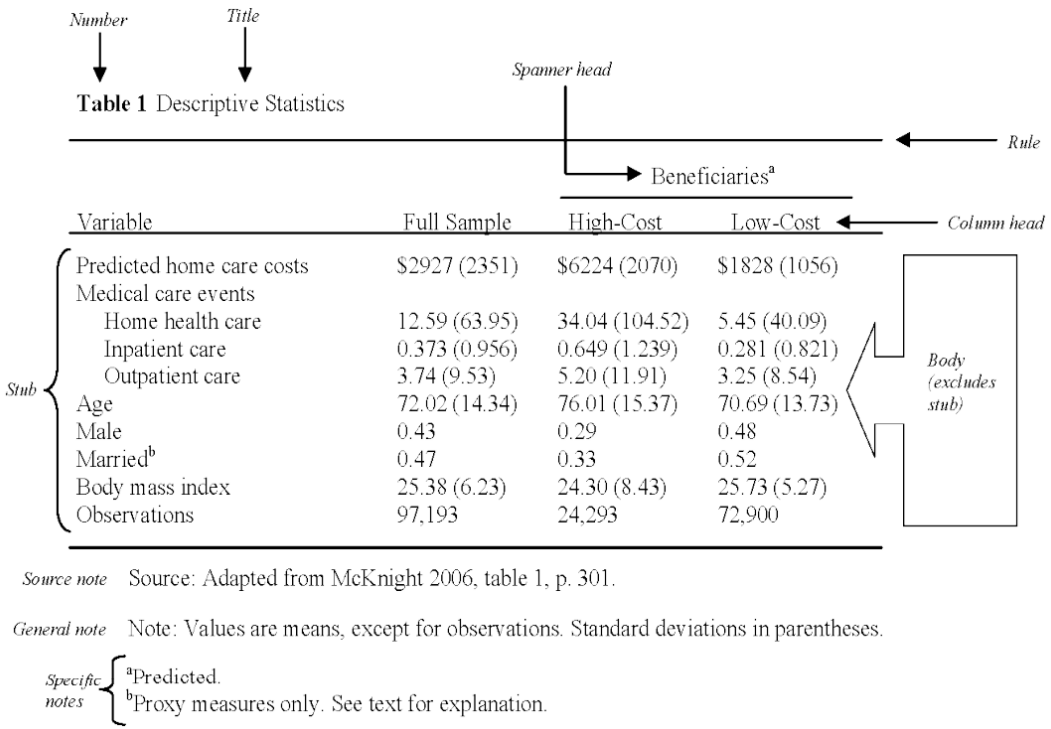
Source note Source: Adapted from McKnight 2006, table 1, p. 301.

General note Note: Values are means, except for observations. Standard deviations in parentheses.

Specific notes [a]Predicted.
[b]Proxy measures only. See text for explanation.

1) It is customary to provide standard deviations, where applicable. In this case, the standard deviations are given in parentheses, right next to the means;

Components of a Table



Appendix: Annotated Table

Here is a sample table of descriptive statistics. The main parts are identified.

Number → **Table 1** Descriptive Statistics ← *Title*

Spanner head → Beneficiaries[a] ← *Rule*

Variable	Full Sample	High-Cost	Low-Cost
Predicted home care costs	$2927 (2351)	$6224 (2070)	$1828 (1056)
Medical care events			
Home health care	12.59 (63.95)	34.04 (104.52)	5.45 (40.09)
Inpatient care	0.373 (0.956)	0.649 (1.239)	0.281 (0.821)
Outpatient care	3.74 (9.53)	5.20 (11.91)	3.25 (8.54)
Age	72.02 (14.34)	76.01 (15.37)	70.69 (13.73)
Male	0.43	0.29	0.48
Married[b]	0.47	0.33	0.52
Body mass index	25.38 (6.23)	24.30 (8.43)	25.73 (5.27)
Observations	97,193	24,293	72,900

Column head → Low-Cost

Stub = Variable column; *Body (excludes stub)* = Full Sample, High-Cost, Low-Cost columns

Source note Source: Adapted from McKnight 2006, table 1, p. 301.

General note Note: Values are means, except for observations. Standard deviations in parentheses.

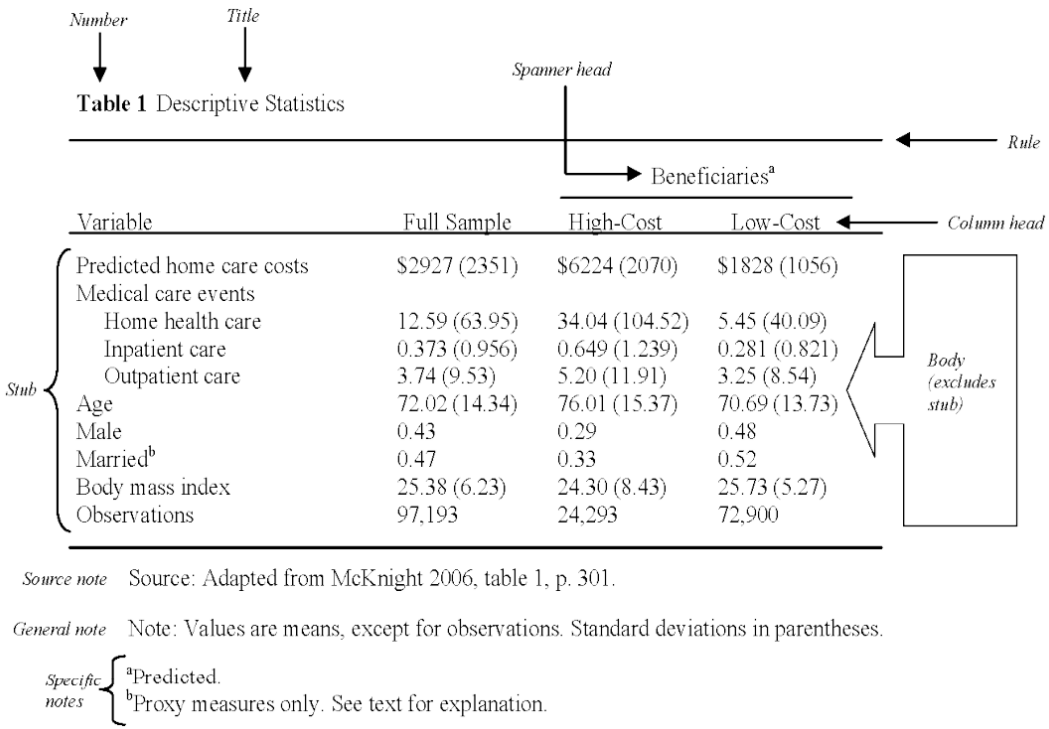
Specific notes [a]Predicted.
[b]Proxy measures only. See text for explanation.

2) The heading "Beneficiaries" is a spanner head because it "spans" or applies across two or more column heads; a spanner rule indicates the relationship between the spanner head and the column heads;

Components of a Table

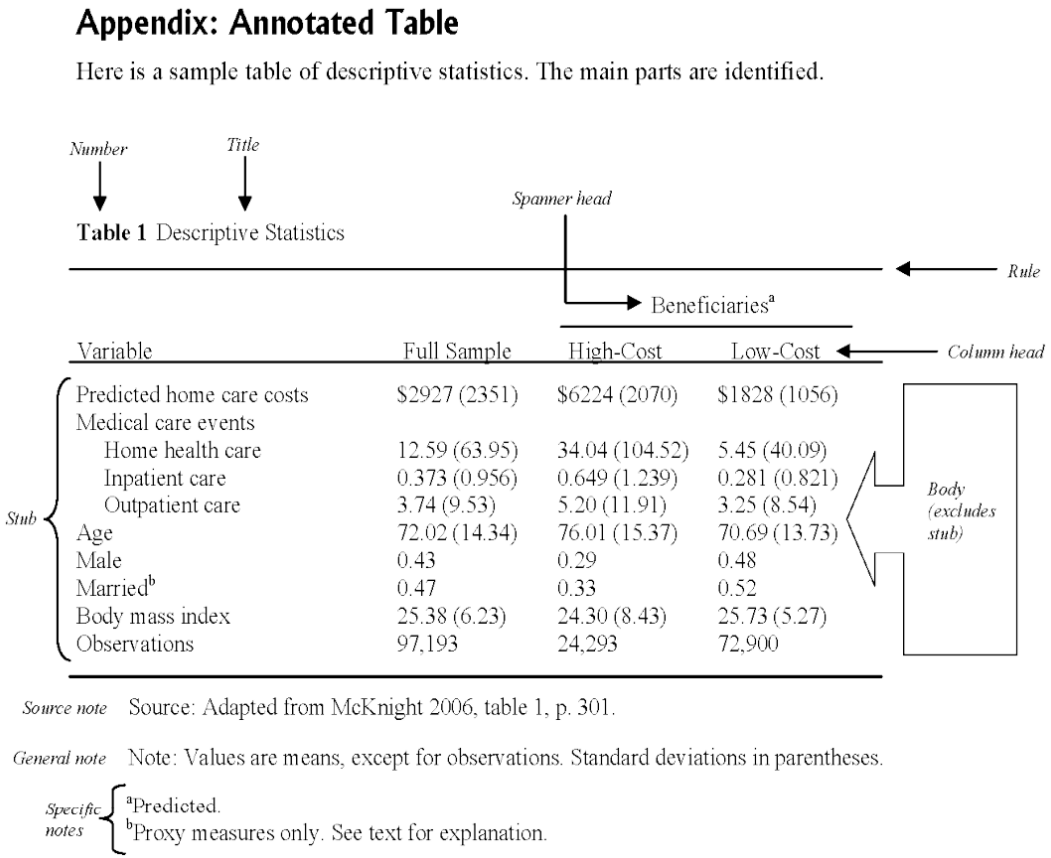
Appendix: Annotated Table

Here is a sample table of descriptive statistics. The main parts are identified.



3) The general note applies to the table as a whole;

Components of a Table



Appendix: Annotated Table

Here is a sample table of descriptive statistics. The main parts are identified.

Number *Title*

Table 1 Descriptive Statistics

Spanner head

Rule

Beneficiaries[a]

Variable	Full Sample	High-Cost	Low-Cost
Predicted home care costs	$2927 (2351)	$6224 (2070)	$1828 (1056)
Medical care events			
Home health care	12.59 (63.95)	34.04 (104.52)	5.45 (40.09)
Inpatient care	0.373 (0.956)	0.649 (1.239)	0.281 (0.821)
Outpatient care	3.74 (9.53)	5.20 (11.91)	3.25 (8.54)
Age	72.02 (14.34)	76.01 (15.37)	70.69 (13.73)
Male	0.43	0.29	0.48
Married[b]	0.47	0.33	0.52
Body mass index	25.38 (6.23)	24.30 (8.43)	25.73 (5.27)
Observations	97,193	24,293	72,900

Column head

Stub

Body (excludes stub)

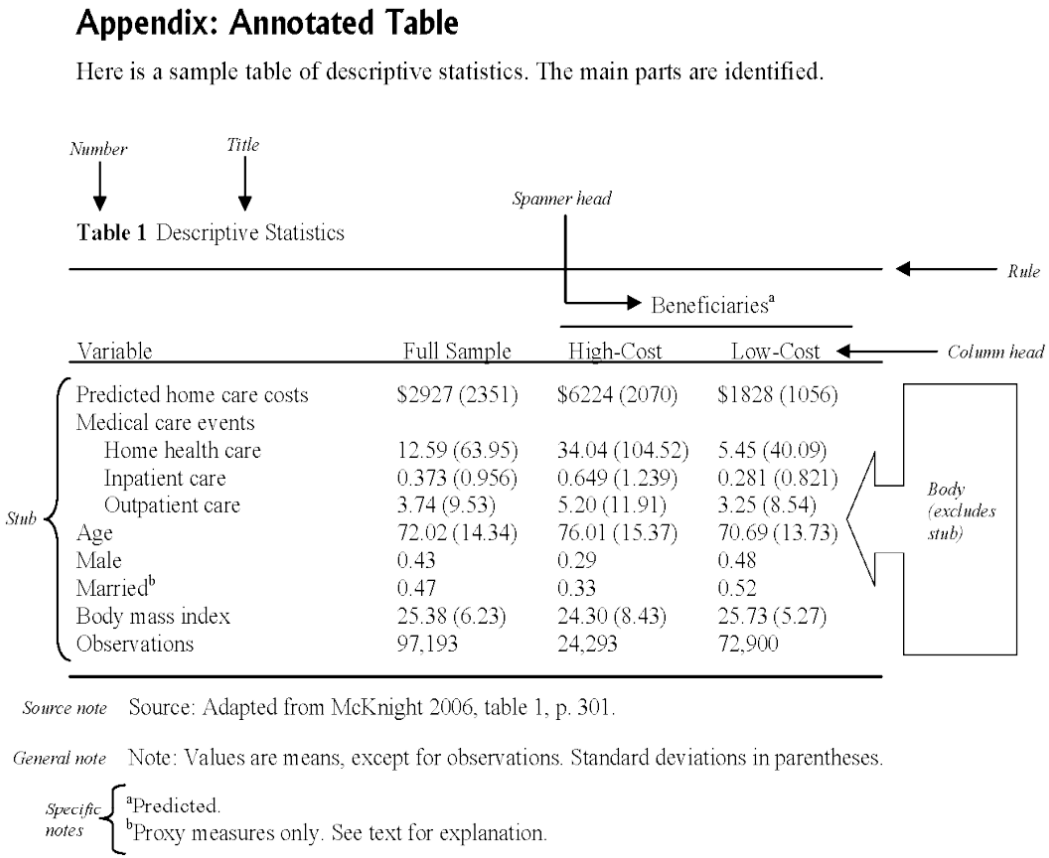
Source note Source: Adapted from McKnight 2006, table 1, p. 301.

General note Note: Values are means, except for observations. Standard deviations in parentheses.

Specific notes [a]Predicted.
[b]Proxy measures only. See text for explanation.

4) The specific note "*a*" applies only to the category "Beneficiaries"; likewise, the specific note "*b*" applies only to the figures for "Married";

Components of a Table



Appendix: Annotated Table

Here is a sample table of descriptive statistics. The main parts are identified.

Number → **Table 1** Descriptive Statistics ← *Title*

Spanner head → Beneficiaries[a]

Rule →

Variable	Full Sample	High-Cost	Low-Cost
Predicted home care costs	$2927 (2351)	$6224 (2070)	$1828 (1056)
Medical care events			
Home health care	12.59 (63.95)	34.04 (104.52)	5.45 (40.09)
Inpatient care	0.373 (0.956)	0.649 (1.239)	0.281 (0.821)
Outpatient care	3.74 (9.53)	5.20 (11.91)	3.25 (8.54)
Age	72.02 (14.34)	76.01 (15.37)	70.69 (13.73)
Male	0.43	0.29	0.48
Married[b]	0.47	0.33	0.52
Body mass index	25.38 (6.23)	24.30 (8.43)	25.73 (5.27)
Observations	97,193	24,293	72,900

Column head ↑

Stub ↓ *Body (excludes stub)* ↑

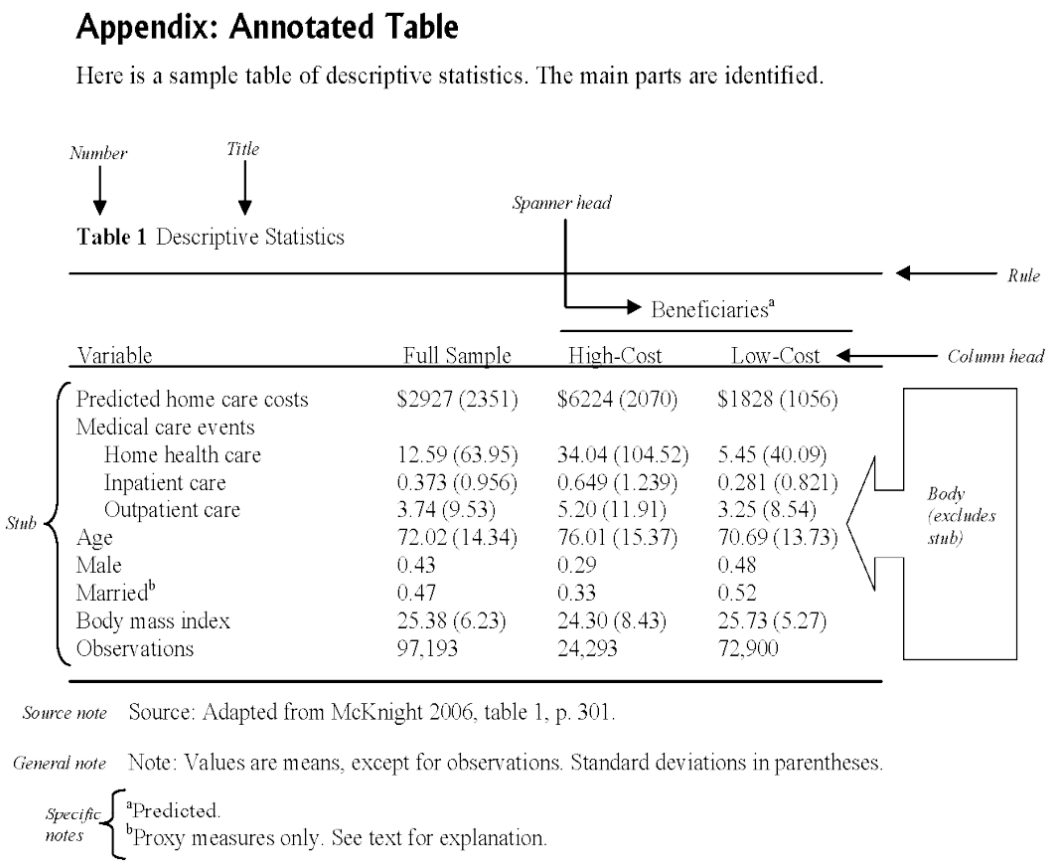
Source note Source: Adapted from McKnight 2006, table 1, p. 301.

General note Note: Values are means, except for observations. Standard deviations in parentheses.

Specific notes [a]Predicted.
[b]Proxy measures only. See text for explanation.

5) Note the absence of **vertical rules**, which are not considered professional and thus should be avoided if at all possible;

Components of a Table



Appendix: Annotated Table

Here is a sample table of descriptive statistics. The main parts are identified.

Number → **Table 1** Descriptive Statistics ← *Title*

Spanner head → Beneficiaries[a]

Rule ←

Variable	Full Sample	High-Cost	Low-Cost
Predicted home care costs	$2927 (2351)	$6224 (2070)	$1828 (1056)
Medical care events			
Home health care	12.59 (63.95)	34.04 (104.52)	5.45 (40.09)
Inpatient care	0.373 (0.956)	0.649 (1.239)	0.281 (0.821)
Outpatient care	3.74 (9.53)	5.20 (11.91)	3.25 (8.54)
Age	72.02 (14.34)	76.01 (15.37)	70.69 (13.73)
Male	0.43	0.29	0.48
Married[b]	0.47	0.33	0.52
Body mass index	25.38 (6.23)	24.30 (8.43)	25.73 (5.27)
Observations	97,193	24,293	72,900

Column head ←

Stub { } *Body (excludes stub)* ←

Source note Source: Adapted from McKnight 2006, table 1, p. 301.

General note Note: Values are means, except for observations. Standard deviations in parentheses.

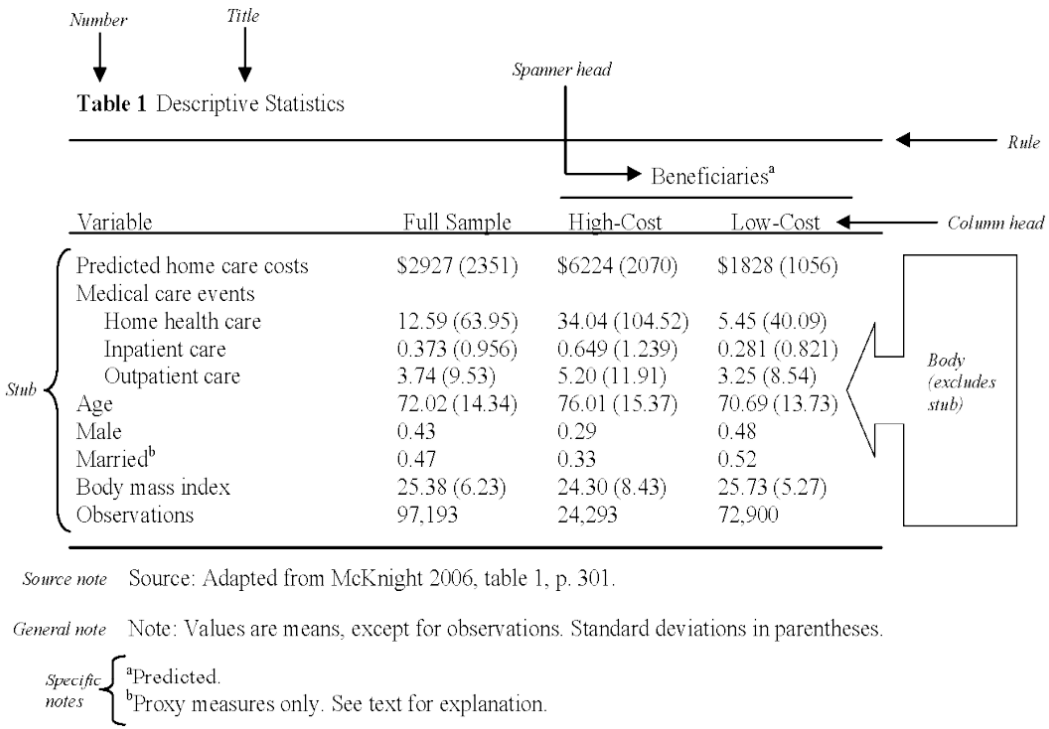
Specific notes [a]Predicted.
[b]Proxy measures only. See text for explanation.

6) In this example, in column heads, all substantive words are capitalized, whereas in the stub entries, only the first word and any proper nouns are capitalized; it is customary to do one or the other: that is, to capitalize all substantive words in the column heads but only first words and proper nouns in the stub, or vice versa;

Components of a Table

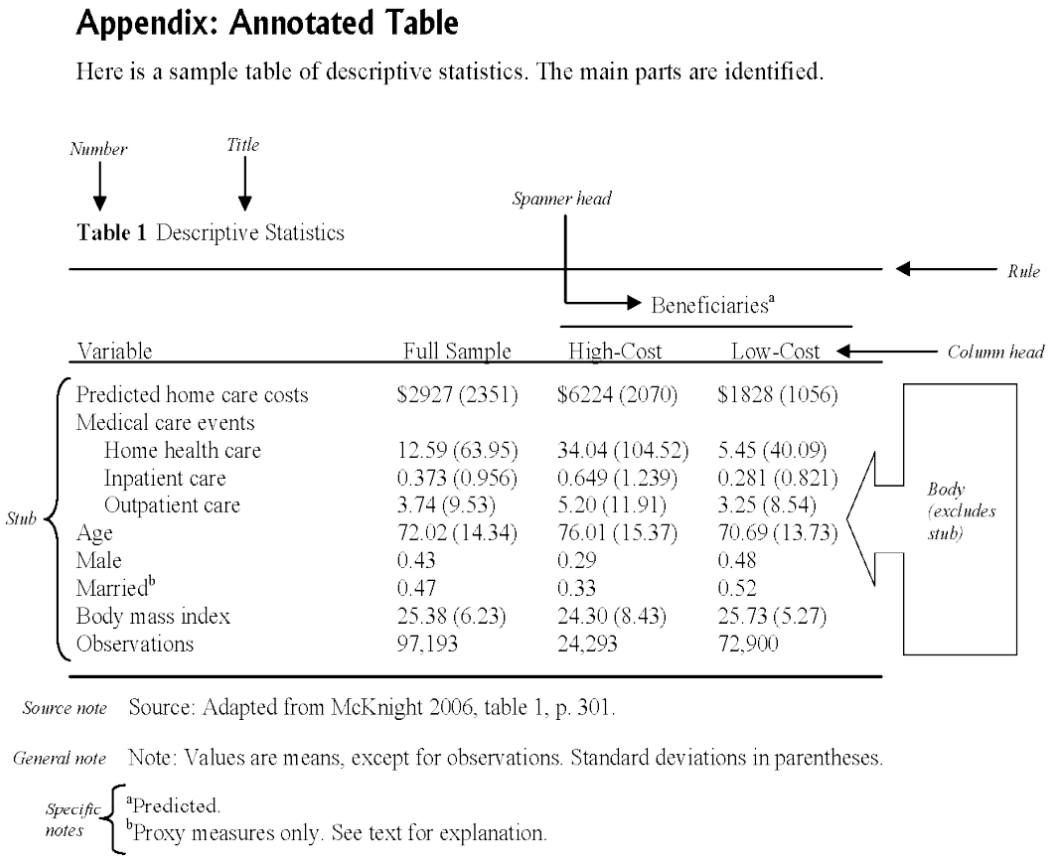
Appendix: Annotated Table

Here is a sample table of descriptive statistics. The main parts are identified.



7) Note that when a stub entry has sub-entries — as in the case of "Medical care events" — the sub-entries are indented;

Components of a Table



Appendix: Annotated Table

Here is a sample table of descriptive statistics. The main parts are identified.

Number *Title*

Table 1 Descriptive Statistics

Spanner head

Rule

Beneficiaries[a]

Variable	Full Sample	High-Cost	Low-Cost
Predicted home care costs	$2927 (2351)	$6224 (2070)	$1828 (1056)
Medical care events			
Home health care	12.59 (63.95)	34.04 (104.52)	5.45 (40.09)
Inpatient care	0.373 (0.956)	0.649 (1.239)	0.281 (0.821)
Outpatient care	3.74 (9.53)	5.20 (11.91)	3.25 (8.54)
Age	72.02 (14.34)	76.01 (15.37)	70.69 (13.73)
Male	0.43	0.29	0.48
Married[b]	0.47	0.33	0.52
Body mass index	25.38 (6.23)	24.30 (8.43)	25.73 (5.27)
Observations	97,193	24,293	72,900

Column head

Stub

Body (excludes stub)

Source note Source: Adapted from McKnight 2006, table 1, p. 301.

General note Note: Values are means, except for observations. Standard deviations in parentheses.

Specific notes [a]Predicted.
[b]Proxy measures only. See text for explanation.

8) The figures in the body of the table should be aligned in some consistent way; they can be aligned by the decimal points; here, they are aligned on the left;

Components of a Table

Appendix: Annotated Table

Here is a sample table of descriptive statistics. The main parts are identified.

Number *Title*

Table 1 Descriptive Statistics

Spanner head

Rule

Beneficiaries[a]

Variable	Full Sample	High-Cost	Low-Cost
Predicted home care costs	$2927 (2351)	$6224 (2070)	$1828 (1056)
Medical care events			
Home health care	12.59 (63.95)	34.04 (104.52)	5.45 (40.09)
Inpatient care	0.373 (0.956)	0.649 (1.239)	0.281 (0.821)
Outpatient care	3.74 (9.53)	5.20 (11.91)	3.25 (8.54)
Age	72.02 (14.34)	76.01 (15.37)	70.69 (13.73)
Male	0.43	0.29	0.48
Married[b]	0.47	0.33	0.52
Body mass index	25.38 (6.23)	24.30 (8.43)	25.73 (5.27)
Observations	97,193	24,293	72,900

Column head

Stub

Body (excludes stub)

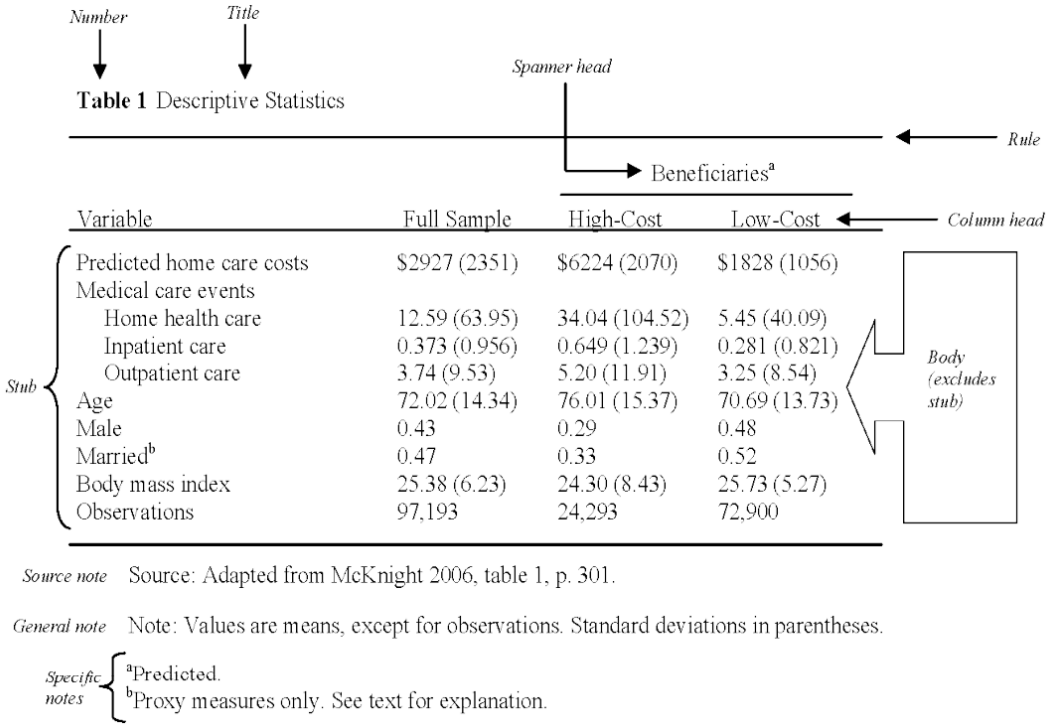
Source note Source: Adapted from McKnight 2006, table 1, p. 301.

General note Note: Values are means, except for observations. Standard deviations in parentheses.

Specific notes [a]Predicted.
[b]Proxy measures only. See text for explanation.

9) The columns heads should be aligned in some consistent way over the columns of figures; here, they are centered;

Components of a Table



Appendix: Annotated Table

Here is a sample table of descriptive statistics. The main parts are identified.

Number *Title*

Table 1 Descriptive Statistics

Spanner head

Rule

Beneficiaries[a]

Variable	Full Sample	High-Cost	Low-Cost
Predicted home care costs	$2927 (2351)	$6224 (2070)	$1828 (1056)
Medical care events			
Home health care	12.59 (63.95)	34.04 (104.52)	5.45 (40.09)
Inpatient care	0.373 (0.956)	0.649 (1.239)	0.281 (0.821)
Outpatient care	3.74 (9.53)	5.20 (11.91)	3.25 (8.54)
Age	72.02 (14.34)	76.01 (15.37)	70.69 (13.73)
Male	0.43	0.29	0.48
Married[b]	0.47	0.33	0.52
Body mass index	25.38 (6.23)	24.30 (8.43)	25.73 (5.27)
Observations	97,193	24,293	72,900

Column head

Stub

Body (excludes stub)

Source note Source: Adapted from McKnight 2006, table 1, p. 301.

General note Note: Values are means, except for observations. Standard deviations in parentheses.

Specific notes [a]Predicted.
[b]Proxy measures only. See text for explanation.

10) The column head for the stub, as well as the main entries in the stub, are always flush left.

Exercise 3

Find the table number, title, column heads (and spanner heads), stub, body, three types of footnotes, and rules (incl. spanner rules) from the table below. How are its contents aligned and indented?

TABLE 9—CROP YIELD, GROWTH CYCLE, AND LONG-TERM ORIENTATION: REGIONAL-LEVEL ANALYSIS BASED ON WVS

	Share of individuals in WVS region with long-term orientation											
	Whole world										Old World	
	Unweighted				Weighted: area				Weighted: area share		Area	Share
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Crop yield	0.049	0.046	0.053		0.097		0.032		0.031		0.039	0.032
	(0.012)	(0.013)	(0.017)		(0.033)		(0.012)		(0.013)		(0.015)	(0.013)
Crop growth cycle			−0.010		−0.047		−0.024		−0.036		−0.027	−0.036
			(0.012)		(0.021)		(0.010)		(0.009)		(0.009)	(0.008)
Crop yield (ancestors)				0.077		0.133		0.043		0.041		
				(0.020)		(0.032)		(0.017)		(0.017)		
Crop growth cycle (ancestors)				−0.012		−0.050		−0.027		−0.037		
				(0.013)		(0.018)		(0.009)		(0.009)		
Continental FE	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	No
Country FE	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Geographical controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Old World sample	No	No	No	No	No	No	No	No	No	No	Yes	Yes
Weighted by region area	No	No	No	No	Yes	Yes	Yes	Yes	No	No	Yes	No
Weighted by region's share of area	No	No	No	No	No	No	No	No	Yes	Yes	No	Yes
Adjusted R^2	0.22	0.25	0.25	0.28	0.28	0.37	0.72	0.72	0.86	0.86	0.72	0.86
Observations	1,356	1,356	1,356	1,356	1,356	1,356	1,356	1,356	1,356	1,356	1,143	1,143

Notes: The table shows the effect of a region's pre-1500CE potential crop yield and its post-1500CE change in the course of the Columbian Exchange (measured in calories per hectare per year) on the share of its population with long-term orientation, accounting of country fixed effects. Geographical controls are absolute latitude, mean elevation above sea level, terrain roughness, percentage of land within 100km of sea, landlocked dummy, and area suitable for agriculture. Columns 1–4 show the unweighted results; columns 5–8 weight observations according to the region's area; columns 9–10 weight observations according to the region's area as a share of the country's area; and columns 11–12 conduct the analysis for the Old World sample. All independent variables have been normalized by subtracting their mean and dividing by their standard deviation. Thus, all coefficients can be compared and show the effect of a one standard deviation increase in the independent variable on long-term orientation. Heteroskedasticity-robust standard error estimates clustered at the country level are reported in parentheses.

Bad Table Design

Common bad table examples include:

- Large and various fonts;
- Too small or too large numbers;
- Badly aligned numbers;
- Colorful and stylish border lines;
- Non-systematic order.

Table Number and Title

Table number:

- Every table should have a number.
- Tables should be numbered consecutively throughout a document.

Table title:

- The title should be brief but descriptive. It should not be a complete sentence, but a collection of words that indicate the subject of the table:
- e.g., "Effect of Class Size on Student Achievement: OLS Regression Results."
- e.g., "Summary of Income Data from Survey in Rural Georgia, 1920–1945."
- Provide estimation method (e.g., OLS estimates), unit of measurement (e.g., US dollars) and period (e.g., 2011-2020) if needed.

Table Column Heads, Stub and Rules

Columns heads:

- Every column of information should have a column head, a word or phrase that identifies the information.
- Spanner heads are used when column heads are in two or more levels, that is, when there are both a collective head and individual heads.

Stub:

- The stub is the very left-most column in a table.
- Do not use variable names used in computer software as labels in stub (e.g., use "Years of education" instead of "edu_yr").

Rules:

- Rules are the lines that visually separate the table into parts.
- In general, only horizontal rules should be used. **Vertical rules** should be avoided.

Table Body

Body:

- The body of a table consists of the columns to the right of the stub and below the column heads.
- A table should have at least two columns and at least six cells of information: two columns and three rows, or three columns and two rows (not including the stub).
- Provide parameter estimates and their standard errors in parenthesis ("()", not "[]" or "{}") in regression tables.

Table Footnotes

Footnotes:

- Three main kinds of footnotes at the end of a table:

  1) Source notes: identify either the source of the data used in the table or, if the table was reproduced without change from a published work. Reproducing a table without change from a published work that is still protected by copyright requires formal permission;

  2) General notes: apply to the table as a whole;

  3) Specific notes: pertain to specific numbers or rows or columns in the table.

Table Style

- Use 10 point *Times New Roman* for labels and 10 point *Courier New* for numbers. Do not use stylish fonts and too big or too small size.
- Rescale numbers appropriately in order to avoid such numbers as "0.00004455" or "75,845,341,697,785."
- Report up to three or four digits below the decimal point. Do not round numbers arbitrarily.
- Do not use stylish border lines (e.g., their colors, thickness, and type of lines).
- Align numbers in a consistent way and consider the location of decimal point carefully.

Exercise 4

What mistakes were made by the author of the table below? Re-design the table in a professional way.

Dependent Variable: Vertical Control

Independent Variables	b	t	Sig.
Constant		21.54	.000
UNCERT	.194	2.231	.027 [a]
TRUST	.098	1.087	.279
TRUST · UNCERT	-.181	-2.049	.042[a]
Adjusted R^2 = .050			

a : Reject Ho at $p < .05$ (1-tailed test)

Exercise 4

Table: The Effects of Uncertainty and Trust on Vertical Control, OLS Estimates

	Vertical control
Uncertainty	0.194**
	(0.087)
Trust	0.098
	(0.09)
Uncertainty X Trust	-0.181**
	(0.088)
Constant	?***
	?
Adj. R-sq	0.050
Observations	?

Note: The table reports OLS estimation results of the effects of uncertainty and trust on vertical control. Robust standard errors are reported in parentheses. *** denotes statistical significance at the 1 percent level and ** at the 5 percent level, all for one-tailed hypothesis tests.

Table of Contents

From Stata to LaTeX/Word/Excel

Most popular Stata packages:
1) estout/esttab
2) outreg2

- Facilitate the convertion of regression outputs to a standard format in publication quality;
- Write LaTeX, MS Word and MS Excel format tables (*estout/esttab* can even export to CSV, HTML, RTF, etc.);
- Report regression outputs, descriptive statistics, frequencies and basic crosstabs.

Exercise 5

Do the Stata exercises in slides:

"Using outreg2 to report regression output, descriptive statistics, frequencies and basic cross-tabulations" (https://dss.princeton.edu/training/Outreg2.pdf).

From Excel to LaTeX

LaTeXTables Generator:

- https://www.tablesgenerator.com/
- *"Easily create even complex LaTeXtables with our online generator – you can paste data from a spreadsheet, merge cells, edit borders and more."*

Exercise 6

Export one of your regression tables generated in Exercise 5 into MS Excel and import it into LaTeXTables Generator to generate a publication-quality LaTeXtable.

Your LaTeXtable is expected to be:

- in "booktabs" table style;
- without vertical rules;
- with texts aligned to the right;
- centered horizontally;
- scaled to text width;
- rotated (in landscape);
- labeled as "tab:exercise";
- captioned above with "The Effects of X on Y, OLS Estimates, 2010-2020".

References

Paul Dudenhefer. "A guide to Writing in Economics". In: *EcoTeach Center and Department of Economics, Duke University* (2009).

Hun Myoung Park. "Practical guides to panel data modeling: a step-by-step analysis using stata". In: *Public Management and Policy Analysis Program, Graduate School of International Relations, International University of Japan* 12 (2011), pp. 1–52.

Oscar Torres-Reyna. "Getting Started in Data Analysis using Stata". In: *Princeton: Princeton University* (2007).