

<http://www.xkcd.com/833>

## STATISTICAL REPORTING SUPPLEMENT to the MBA Mandatory Writing Guide

## OVERVIEW & PURPOSE OF THIS GUIDE

This is a supplemental guide to the reporting of the results of statistical analyses for a general management or other non-technical audience. The emphasis of this guide is the interpretability and objective accuracy of such reports. The precept governing the construction of this guide is that statistics are tools and not answers, and that the use of statistical analysis in the context of a business report should be subordinate to and supportive of the clear exposition of the case to be argued or finding to be explained. To that end, this guide will present style and formatting rules for both the in-line reporting of statistical results within a narrative and the construction of stand-alone tables, charts, or figures that present detailed statistics to clarify the narrative. In large part, this guide is aligned with APA Style reporting standards codified in the *Publication Manual of the American Psychological Association*, 6th edition, Washington, DC: APA, 2010.

## HOW TO READ & USE THIS REPORTING GUIDE

This guide is roughly formatted to follow the progression of a typical foundation-level statistics course, with basic descriptive and visual statistical techniques covered first and increasingly more complex inferential and analytical techniques to follow. Basic statistical notation and symbol usage is also covered. Although the guide will suffice to provide rules for the accurate and concise treatment of statistical *results* in business writing, the stylistic subtleties involved in the accurate portrayal of *information* and *conclusions* based on statistical results are too broad a topic to cover in brief. The standard style sourcebook where the reporting of statistical data and conclusions is concerned is *The Chicago Guide to Writing about Numbers*, Jane E. Miller, Chicago, IL: Univ. of Chicago Press, 2004. This book is a must for anyone who uses data often.

The formatting used in this guide includes **highlighted text** to indicate specific writing examples and in-line figures or tables to demonstrate the appropriate use of such reporting tools. The major **BOLDFACE HEADINGS** in the guide correspond to broad areas of statistical analysis and the associated reporting styles, with specific concerns connected to each area being outlined or headed by Paragraph Headings within the body of each section where such headings are useful for clarity or quick reference. Very often, there are two or more ways of reporting the same information with equal accuracy, the choice of style being dependent on the context and purpose of the statistical result to be conveyed. Guidelines are given in such cases to help in choosing the most appropriate reporting style for the purpose and the intended audience.

## STATISTICAL REPORTING IN A BUSINESS CONTEXT

Statistical data and analysis are the soul of sound business decision-making. The goal of any statistical report in business should be to inform decision-making. For this reason, it is critical for a statistician or analyst to be aware of the external environment and internal competencies that impinge on and relate to any analysis to be conducted. In a business setting, any statistical report that does not take into account such factors should be considered incomplete and must be reworked. An acceptable statistical report in business has relevant, actionable conclusions.

## GENERAL RULES OF STATISTICAL REPORTING

### Reporting Structure

A good statistical report is intentional and well-focused, beginning with a research question or business problem to be answered, continuing to an explanation of the data and analytical tools to be used, presenting a context for interpretation that includes both descriptive and inferential statistical results, and ending with a well-defended conclusion. Details of data transformation or cleaning, meta-analytic statistics (*i.e.*, fit, post-hocs), or similar usually belong in an appendix.

### Rounding & Precision

Except in situations where great precision is warranted or an exact  $p$ -value is desired, statistics should be reported rounded to two decimal places. When this is not practical because the scale of measurement for a variable requires multiple leading zeroes—*e.g.*, **the test procedure had an error rate of 0.0064**—the variable may be rescaled or the rule amended to two significant digits.

### Reporting Percentages

Percentages, proportions, or probabilities should be written out if intended to be read within the narrative—*e.g.*, **less than 40 percent of those surveyed responded with “very satisfied”**. Percentages that are referenced as a numerical quantity should be reported as decimal values—*e.g.*, **the percentage response rate was 0.56**. The percentage symbol, “%”, should be avoided.

### Statistics in the Narrative

If a statistic is to serve as a part of speech in a sentence within the narrative, the proper name of the statistic should be spelled out to whatever extent possible—*e.g.*, **the sample average was 4.65 years with a standard error of 0.41 years**. Any units of measurement or accounting should be given when a statistical value is reported—*e.g.*,  **$x = 505.21$  parts,  $s_x = 25.14$  parts**. In general, most statistics should be reported either within the narrative itself or in a table, but not both.

### Statistical Symbols & Notation

Statistics may be referred to by using the appropriate mathematical symbol or Greek letter, or by using commonly accepted abbreviations in Roman letters. When Roman letters are used, the reference should be italicized—*e.g.*,  **$M$ ,  $SD$ ,  $t(df)$** . When statistical notation is used that requires exponents or subscripts, appropriate font formatting should be employed to this end—*e.g.*,  **$R^2$ ,  $s_x$ ,  $\sigma_x^2$** . When mathematical symbols are used to refer to their associated statistics, population parameters should be reported either with uppercase Roman letters or the appropriate Greek letter—*e.g.*,  **$\mu_x$ ,  $\sigma_x^2$ ,  $N$ ,  $\beta_1$** —and all sample statistics should be reported with lowercase Roman letters—*e.g.*,  **$x$ ,  $s_x^2$ ,  $n$ ,  $b_1$** . An uppercase  **$N$**  may also be used to refer to an overall sample size when sub-groups exist, in which case a lowercase  **$n$**  should be used to refer to the size of each group individually. When abbreviations are used, all letters should be uppercase—*e.g.*,  **$M$ ,  $SD$** .

### Statistical Calculations & Equations

In general, it is not necessary to show or discuss how statistics are calculated. If it is desirable to show a complex or unfamiliar equation, the equation should be created using equation design tools, as available, within the word processor software that is being used—*e.g.*,  **$z = \frac{x - \mu_0}{\sigma_x / \sqrt{n}}$** .

## REPORTING DESCRIPTIVE OR SUMMARY STATISTICS

At a minimum, any statistical reporting should include the mean and standard deviation of the data to be analyzed. If the analysis to be conducted involves statistical inference from a sample, the standard error of the mean may supplement or substitute for the standard deviation. In addition, the sample size should be reported, usually with the first mention of a given sample.

In the body of a report, summary statistics may be reported directly within the narrative—*e.g.*, *the average age within the sample was 25.78 years*—or reported within parentheses following a general description of the sample—*e.g.*, *the sample consisted of undergraduate students, and so had a relatively low average age ( $M = 19.54$  years,  $SD = 1.33$  years,  $n = 250$ )*. These methods for reporting summary statistics may be used simultaneously—*e.g.*, *the average age of graduate students sampled ( $n = 85$ ) was 26.33 years ( $SD = 4.20$  years)*. When summary statistics are given at the end of a sentence, the parentheses may be omitted and a comma used to introduce the results—*e.g.*, *a total of 458 student responses were collected,  $M = 23.55$  years,  $SD = 5.65$  years.*

### Tables of Summary Statistics

When reported in a table, summary statistics are most often given as the columns of the table, with the sample(s) or variable(s) to be summarized given as the rows. The sample size may be given as its own column in the table, or may be indicated in the descriptive title of the table. The headers given to the summary statistic columns may be the proper name of the statistic reported in the column, but are more often mathematical symbols or standard abbreviations. Tables should be formatted without vertical separator lines. Decimal points should be aligned. Units of measurement or accounting are not required to be given when summary statistics are reported in a table. If necessary for clarity, units may be given in the row label for each variable.

Example Table 1

*GMAT Scores of Male and Female MBA Applicants*

	<i>Sample Size</i>	<i>Mean</i>	<i>Std. Deviation</i>
Male	120	554.3	75.3
Female	140	530.2	60.2
Total	260	541.1	67.2

Example Table 2

*Other Admission Factors ( $N = 260$ )*

	<i>M</i>	<i>SD</i>
Work (Years)	5.63	2.25
UGrad GPA	3.11	0.36
GRE Score	480.2	38.5

### Complex Tables of Summary Statistics

When the desired presentation of summary statistics involves multiple samples and multiple variables, the groups or samples to be summarized may be given as the columns of the table with the variables of interest given as the rows. The sample size should be given in parentheses beneath the column label for each group. The cells of the table should contain the calculated mean for each intersection in the cross-tabulation, as well as the associated standard deviation given in parentheses below the reported mean. In order to distinguish from a cross-tabulation or contingency table, the description should indicate that the values reported in the table are averages. It is not strictly necessary to identify the standard deviations, but a note is helpful.

Example Table 3

*Mean Admission Characteristics of MBA Applicants by Undergraduate Major*

	<i>Business</i> ( <i>n</i> = 115)	<i>Science</i> ( <i>n</i> = 58)	<i>Humanities</i> ( <i>n</i> = 72)	<i>Other</i> ( <i>n</i> = 15)
GMAT Score	538.7 (70.21)	560.2 (60.35)	522.8 (65.46)	573.5 (78.44)
GRE Score	476.7 (40.35)	495.8 (34.68)	462.6 (37.62)	531.2 (43.31)
UGrad GPA	3.25 (0.40)	3.30 (0.27)	2.81 (0.40)	3.35 (0.40)
Work (Years)	4.64 (1.89)	7.21 (2.78)	6.42 (2.55)	3.32 (1.52)

*Note: Standard deviations are reported in parentheses below reported means*Cross-tabulations & Contingency Tables

When a frequency distribution of data across multiple categorical variables is to be reported in a cross-tabulation (*a.k.a.*, a contingency table), such as in preparation for a chi-square test, the levels or values taken by each categorical variable should be clearly indicated across the column and row headers as appropriate. If more than two variables are to be summarized, the levels or values of two or more variables may be given side-by-side across the column or row headers or may be used as super-column or super-row headers. A legend of table entries should be given.

Example Table 4

*Distribution of MBA Applicants by Undergraduate Major, Gender, and School Type*

<i>Gender</i>	<i>School</i>		<i>Business</i>	<i>Science</i>	<i>Humanities</i>	<i>Other</i>	<i>Total</i>
Male	Public	Count (#)	29	16	5	4	54
		Expected #	27.5	15.1	9.7	2.2	---
		% Parent Row	0.54	0.29	0.09	0.07	0.45
	Private	Count (#)	32	17	16	1	66
		Expected #	33.7	18.5	11.9	2.6	---
		% Parent Row	0.48	0.26	0.24	0.02	0.55
Female	Public	Count (#)	26	9	18	3	56
		Expected #	21.8	10.1	20.2	3.9	---
		% Parent Row	0.46	0.16	0.32	0.05	0.40
	Private	Count (#)	28	16	33	7	84
		Expected #	32.8	15.1	30.2	5.9	---
		% Parent Row	0.33	0.19	0.39	0.08	0.60
Total	Subtotal	Count (#)	61	33	21	5	120
		Expected #	52.8	26.4	33.6	7.2	---
		% Parent Row	0.51	0.28	0.18	0.04	---
Total	Subtotal	Count (#)	54	25	51	10	140
		Expected #	61.6	30.8	39.2	8.4	---
		% Parent Row	0.39	0.18	0.36	0.07	---
Total	Subtotal	Count (#)	115	58	72	15	260
		Expected #	115	58	72	15	---
		% Parent Row	0.44	0.22	0.28	0.06	---

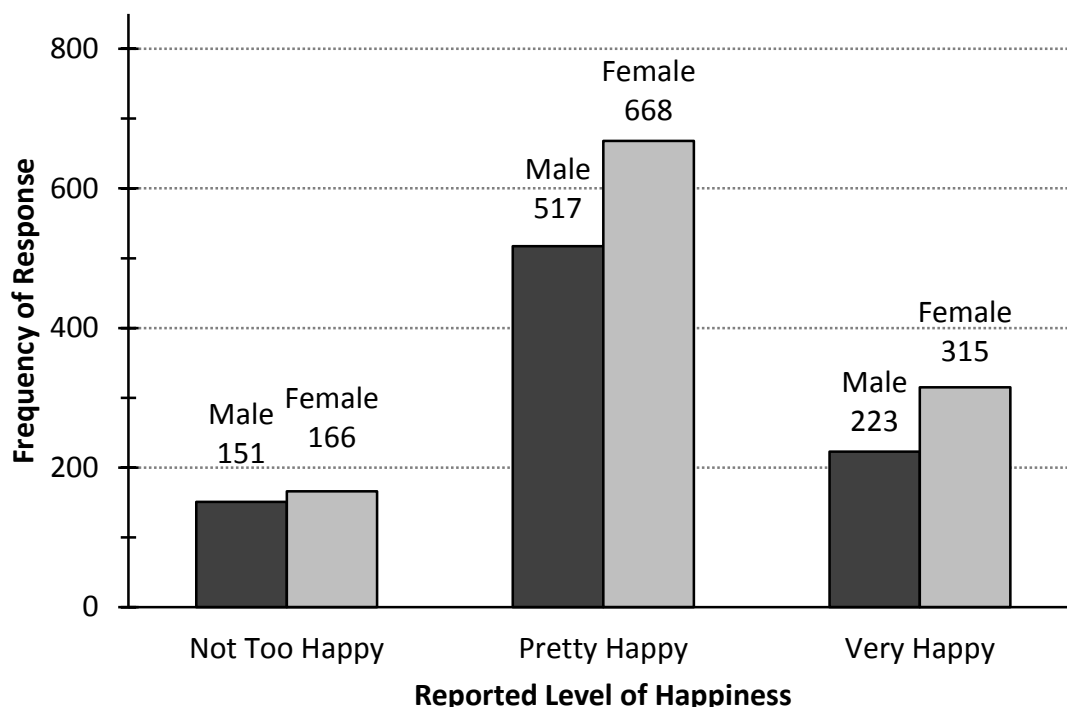
## REPORTING STATISTICS OR DATA IN CHARTS OR FIGURES

When reported graphically in a chart or other figure, summaries or distributions of categorical or continuous variables should be presented with a high data-ink ratio in clean, unornamented graphical figures (c.f., Edward Tufte and P. R. Graves-Morris, *The Visual Display of Quantitative Information*, Volume 31, Cheshire, CT: Graphics Press, 1983). In general, legends should not be used, with preference given to data tables for complex structures or to data labels for simpler structures. All axes must be labeled clearly. The range of the Y axis should extend above and below the range of observed data points and nearly always to a minimum of 0 at the origin. In general, gridlines should be removed or minimized. No border or background color should be applied to either the plot area or the figure itself. The X:Y aspect ratio of the plot area should be approximately between 4:3 and 5:3 in order to minimize any distortion of sizes or relationships.

If the data are categorical, a bar chart should be used in preference over a line chart, and the bars should be separated by gaps. A pie chart may be used to summarize categorical data with no more than five categories, but a pie chart should be avoided if the frequency distribution is relatively uniform or if multiple categorical distributions are to be compared to each other. If the data are continuous (*i.e.*, a histogram), either a bar chart or a line chart may be used. An area chart may also be used as a substitute or extension of a line chart. If a bar chart is used to display a histogram, the bars should not be separated by gaps. If multiple variables or groups are to be summarized, the decision to present the data in a single chart should be purposeful. In general, no more than five variables should be summarized on a single chart, and no more than five categories or histogram bins should be included when summarizing multiple variables.

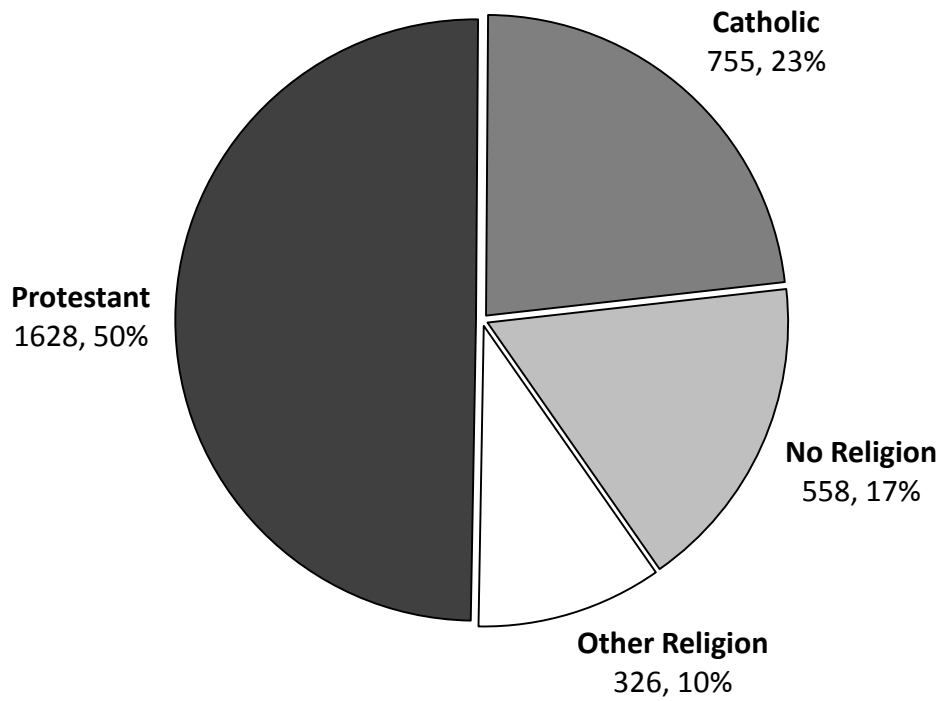
Example Figure 1

*Happiness of Male and Female Respondents (N = 2040)*



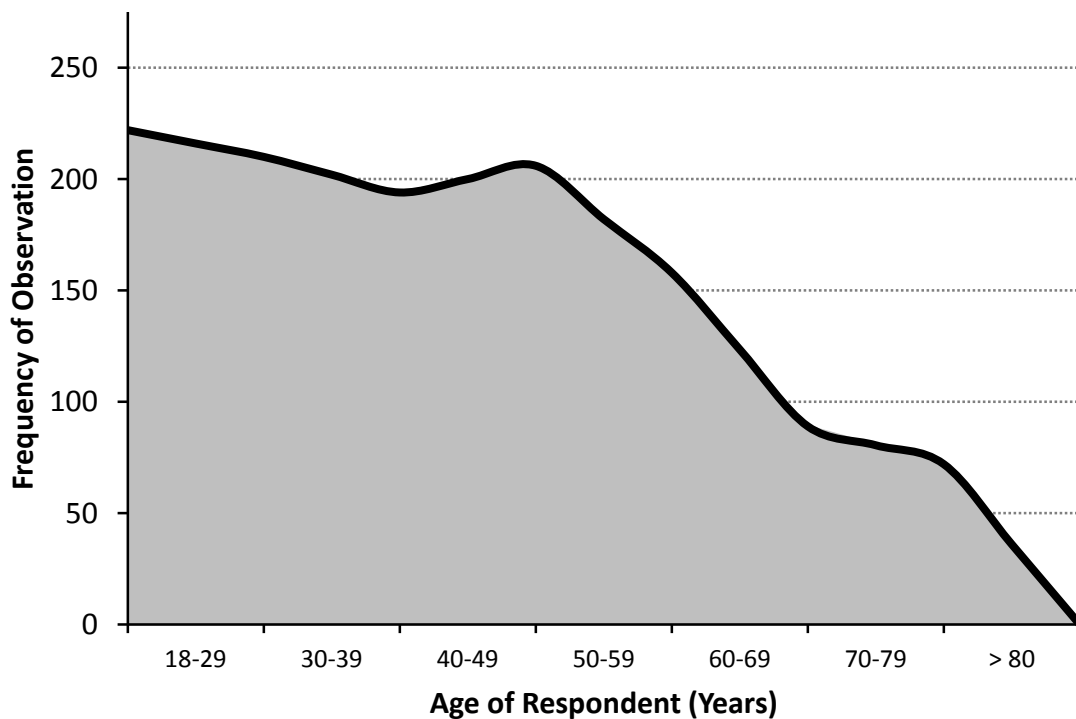
Example Figure 2

*Distribution of Religious Preferences of Respondents (N = 3267)*



Example Figure 3

*Distribution of Ages for Female Respondents (n = 1151)*



## REPORTING SAMPLE ESTIMATES & CONFIDENCE INTERVALS

The most basic use of inferential statistics is the construction of confidence intervals to surround a point estimate of a population parameter. Regardless of the method to be used in reporting a confidence interval, the confidence level should always be indicated clearly as a percentage. The estimated confidence interval itself may be reported either as a point estimate accompanied by the margin of error, or as the actual calculated upper and lower limits of the confidence interval.

In the body of a report, confidence intervals may be reported directly within the narrative—*e.g.*, **the 95% confidence interval for the true proportion of New Orleans residents who do not have health insurance was 0.157 to 0.272**—or reported within brackets accompanying the reported mean and standard deviation or error—*e.g.*, **among residents of the Lower Ninth Ward ( $n = 200$ ), the estimated proportion without health insurance was 0.251,  $SE = 0.031$ , 95% CI [0.191, 0.311]**. The confidence interval may also be reported in terms of a margin of error—*e.g.*, **the proportion of New Orleans residents concerned about medical bills was estimated as  $0.35 \pm 0.07$  (CL = 95%)**.

When reported in a table, confidence intervals should be reported as a lower and upper limit in two adjacent columns of the table with the confidence level clearly indicated in a heading that spans both columns. All other rules concerning the construction of tables should be observed.

Example Table 5

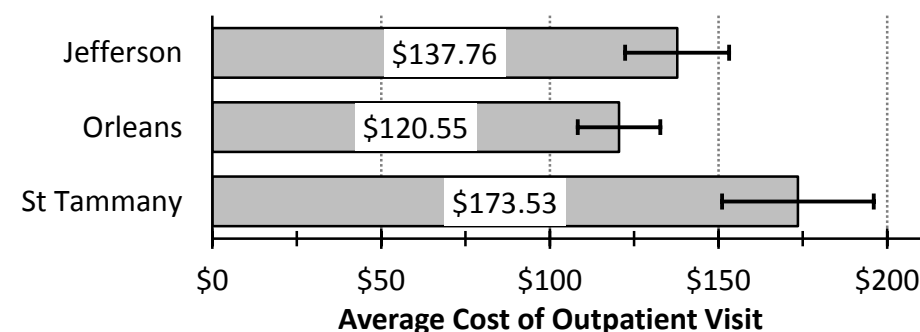
*Estimated Monthly Health Insurance Premiums in New Orleans*

	Sample Size	Mean	Std. Error	99% Confidence Interval	
				Lower Limit	Upper Limit
Family	750	\$335.35	\$6.03	\$319.78	\$350.93
Individual	500	\$142.28	\$2.39	\$136.10	\$148.45

When reported graphically, confidence intervals are shown as a pair of error bars around a point estimate. When this method is used, a note should be given to clarify whether the length of the error bars represents only the standard error or the full interval, as either choice is acceptable.

Example Figure 4

*Estimated Cost of Outpatient Care in New Orleans and Surrounding Parishes*



Excel's built-in chart formatting tools have the ability to create custom error bars.

This function was used to create the figure shown, using a defined cell range for the interval sizes.

*Note: Error bars represent 95% confidence interval of estimated means*



## REPORTING INFERENCE TEST STATISTICS & RESULTS

In addition to the generally required summary statistics, the two required elements in reporting the result of any statistical hypothesis test, regardless of the test procedure that was used, are the calculated test statistic and the  $p$ -value. The test statistic and the associated  $p$ -value should either be reported clearly within the body of the report or in a table summarizing the result of the hypothesis test. In addition, unless the hypothesis test was conducted using  $z$ -scores from a population with known variance, it is required to report the degrees of freedom associated with the calculated test statistic. It is not necessary to describe the procedure that was used, only the result of the test and the conclusion that was reached. If the audience is not likely to be familiar with the test procedure and needs to understand the process, a brief description may be given.

## REPORTING FORMAL HYPOTHESIS STATEMENTS

In general, formal hypothesis statements need not be reported in a general audience summary of the result of a statistical hypothesis test. The research question or problem statement should be reported at least implicitly, and suffices as a description of the test in most cases. When it is desirable to give the formal hypothesis statements, the hypotheses should be presented either in mathematical equations concerning estimated parameters or in concise testable statements.

### Example Hypothesis 1

*Test of Mean Differences in Day and Night Shift Packing Errors*

$H_0: \mu_{\text{Day}} - \mu_{\text{Night}} = 0$  packing errors

$H_A: \mu_{\text{Day}} - \mu_{\text{Night}} \neq 0$  packing errors

### Example Hypothesis 2

*Chi-Square Test of Independence between Job Performance Rating and Shift*

$H_0$ : Job performance ratings are independent of work shift

$H_A$ : Job performance ratings are dependent on work shift

## REPORTING CONCLUSIONS & GENERAL RESULTS

Whether statistical test results are reported within the narrative itself or in a table, the essential conclusion of a statistical hypothesis test should be described in the body of the report in clear language that references the significance and directionality of the result. If a results table is not given, the test statistic and  $p$ -value should follow the conclusion—*e.g.*, **the mean labeling error rate was found to be significantly greater than the target rate of 3 per 10000,  $z = 2.13$ ,  $p = 0.017$ .**  $P$ -values should usually be given rounded to three decimal places, counter to previous guidelines for reporting decimals. An alternative method is to report the highest level of significance that a test could have met—*e.g.*,  $p < 0.001$ ,  $p < 0.01$ ,  $p < 0.05$ . When test statistics are given in a table, this is achieved with asterisks explained in a note—*e.g.*, \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

## REPORTING 1- & 2-SAMPLE TESTS OF MEANS

Basic statistical hypothesis tests that involve comparison of an estimated mean against either a hypothesized value (*i.e.*, a 1-sample test) or another estimated mean (*i.e.*, a 2-sample test) are assumed to be two-tailed in the absence of explicit annotation to the contrary. A one-tailed test should be described clearly within the narrative as having a directional specification, and all one-tailed  $p$ -values should be explicitly noted. If not given either in the preceding narrative or in a summary table, the estimated mean(s) and standard deviation(s) or standard error(s) should be included as a parenthetical in the description of the test result and reporting of test values.

### 1-Sample Test of Mean Values

If the test was conducted using  $z$ -scores from a population with known variance, the reported result should include the calculated test statistic and the associated  $p$ -value—*e.g.*, **the average net fill weight of dark roast coffee ( $n = 60$ ,  $\bar{x} = 1.992$  pounds,  $\sigma = 0.0475$  pounds) was not found to be significantly less than the labeled net weight of 2 pounds,  $z = -1.30$ ,  $p = 0.096$  (one-tailed).**

If the test was conducted using Student's  $t$ -scores in sampling from a population with unknown variance, the test statistic should be reported with the degrees of freedom associated with the observed sample size (or other conditions of the test) indicated in parentheses attached to the  $t$  designator—*e.g.* **the acidity of brewed dark roast coffee ( $n = 14$ ,  $\bar{x} = \text{pH } 5.54$ ,  $s_x = \text{pH } 0.25$ ) was found to be significantly different from the desired acidity of pH 5.69,  $t(13) = -2.24$ ,  $p = 0.043$ .**

### 2-Sample Test of Mean Differences

The procedure for reporting the result of a 2-sample test of mean differences is essentially the same as for a 1-sample test against a hypothesized mean. The area of concern that is particular to 2-sample reporting is one of practicality rather than technical formatting. Reporting two sets of summary statistics, as well as associated test values, within a single sentence is not ideal for clarity or for readability. It is recommended to report a table of summary statistics separately.

Example Table 6

*Mean Coffee Consumption (Cups per Week) by Preference for Dark or Medium Roast*

	Sample Size	Mean	Std. Error	95% Confidence Interval	
				Lower Limit	Upper Limit
Dark Roast	40	3.45	1.11	3.095	3.805
Medium Roast	40	3.00	0.80	2.744	3.256

Given summary statistics reported as in Example Table 6, a concise and accurate conclusion may be written without a need to reference every statistic for each sample. It is nevertheless recommended to repeat the two means when reporting the outcome of the test. This may be accomplished in the narrative itself rather than with parenthetical references as would be used for a single sample—*e.g.*, **the average of 3.45 cups of coffee consumed per week by customers who prefer dark roast was found to be significantly different from the average of 3.00 cups of coffee consumed per week by customers who prefer medium roast,  $t(71) = 2.08$ ,  $p = 0.041$ .**

## REPORTING ANOVA FOR MULTIPLE-GROUP TESTS

Unlike the result of a 1-sample or 2-sample test, which can be reported in a single sentence or summary table, the result of an ANOVA for multiple-group mean comparisons usually requires reporting two or three tables: 1) a table of estimated summary statistics; 2) a table summarizing the results of the ANOVA; and, as needed, 3) a table describing the result of any post-hoc tests.

These tables should be presented in tandem with a written interpretation of the test result. The summary statistics table should follow the previous guidelines for such tables. In this case only, due to the ANOVA method, it is acceptable to substitute the variance for the standard deviation or standard error. The ANOVA result table should include at least the sum of squares, degrees of freedom, mean square, and the  $F$  test statistic. The  $p$ -value may be included or given in a note.

Example Table 7

*Mean Cyclist Weight (lbs.) by Category of Competition*

	Sample Size	Mean	Variance	95% Confidence Interval	
				Lower Limit	Upper Limit
Track Cyclists	20	156.85	328.10	148.37	165.33
BMX Riders	18	169.67	383.62	159.93	179.41
Mountain Bikers	17	145.76	325.48	136.49	155.04

Example Table 8

*Summary of ANOVA for Mean Cyclist Weight Comparison*

	Sum of Squares	df	Mean Square	F
Between Groups	5013.37	2	2506.69	7.26**
Within Groups	17963.27	52	345.45	
Overall	22976.64	54		

Note: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Given summary statistics and ANOVA results reported as in Example Tables 7 and 8, a written conclusion may be drafted with reference to the tables as needed. In reporting the calculated test statistic, both the numerator and the denominator degrees of freedom must be indicated in parentheses attached to the  $F$  designator, with numerator  $df$  given first and denominator  $df$  separated by a comma—e.g., *in sampling cyclists in all three categories of Olympic competition, significant differences were found in average cyclist weight by category,  $F(2, 52) = 7.26, p < 0.01$ .*

If desired, a calculated measure of effect size such as  $\eta^2$  (eta squared) or  $\omega^2$  (omega squared) may be reported; such a measure must be clearly narrated and identified—e.g., *the category of competition was found to explain 21.8 percent of the variance in cyclists' weights ( $\eta^2 = 0.218$ ).*

### Post-Hoc Testing for ANOVA

If any post-hoc tests were conducted for pair-wise mean differences, the results of the testing should be reported in a third table. Any of the common post-hoc tests (including Fisher Least Significant Differences, Tukey Honest Significant Differences, and Bonferroni Corrections) may be reported using a single table format. The test used should be identified clearly in the title.

The essential components of a table presenting post-hoc test results for ANOVA are the mean pair-wise difference between each pairing of groups, the pooled standard error of each mean difference, and an indication of the significance of each mean difference. A confidence interval for each mean difference may also be included. The group pairings may be listed with each pairing in a single row or with pairings given as a cross-tabulation spanning two columns.

Example Table 9

*Bonferroni Corrected Comparisons of Cyclist Weights (lbs.) between Competitor Groups*

				<i>95% Confidence Interval</i>	
		<i>Mean Diff.</i>	<i>Std. Error.</i>	<i>Lower Limit</i>	<i>Upper Limit</i>
BMX Riders	vs. Mountain Bikers	23.91**	6.38	10.93	36.87
BMX Riders	vs. Track Cyclists	12.82	6.12	0.41	25.22
Track Cyclists	vs. Mountain Bikers	11.09	5.96	-1.02	23.19

Note: \* Bonferroni corrected  $p < 0.05$ , \*\* Bonferroni corrected  $p < 0.01$

Example Table 10

*Tukey HSD Comparisons of Cyclist Weights (lbs.) between Competitor Groups*

Group 1	Group 2	Mean Diff.	Std. Error.	95% Confidence Interval	
				Lower Limit	Upper Limit
Track Cyclists	BMX Riders	-12.82	6.12	-25.22	-0.41
	Mountain Bikers	11.09	5.96	-1.02	23.19
BMX Riders	Track Cyclists	12.82	6.12	0.41	25.22
	Mountain Bikers	23.91**	6.38	10.93	36.87
Mountain Bikers	Track Cyclists	-11.09	5.96	-23.19	1.02
	BMX Riders	-23.91**	6.38	-36.87	-10.93

Note: \*  $p < 0.05$  (Tukey HSD = 14.85 lbs.), \*\*  $p < 0.01$  (Tukey HSD = 19.17 lbs.)

After post-hoc testing is reported in a table as above, a written summary of the analysis should be provided with reference to the post-hoc, prior summary statistics, or ANOVA results tables as needed—e.g., **post-hoc analysis of pair-wise group comparisons using the Tukey HSD criteria indicated that only the mean difference between the BMX Rider ( $M = 169.67$  lbs.,  $SE = 4.62$  lbs.) and Mountain Biker ( $M = 145.76$  lbs.,  $SE = 4.38$  lbs.) competitor groups was significant,  $p < 0.01$ .**

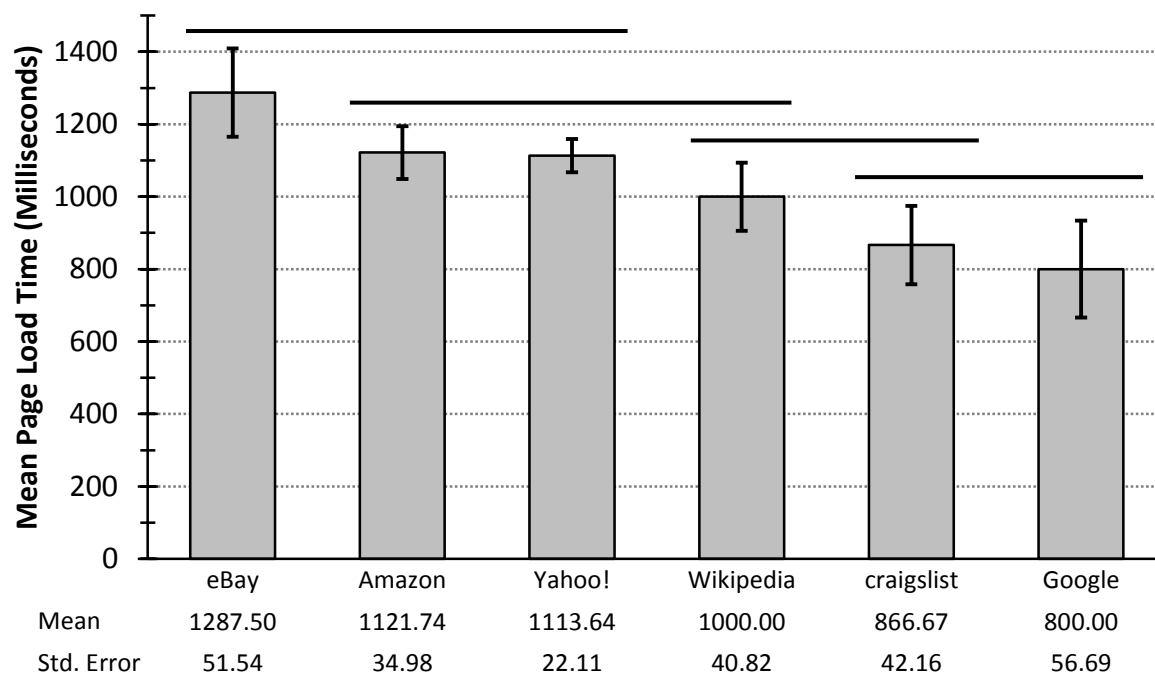
## REPORTING MEAN COMPARISONS GRAPHICALLY

When a comparison of means conducted with a t-test or ANOVA is reported graphically, the guidelines for the general preparation of charts and figure should be observed. Additionally, new methods for labeling or organizing the chart or figure are needed in order to indicate the result of each mean comparison relative to the assigning of statistical significance to the mean difference. There are two primary methods: 1) to graph the various group means as usual and apply additional labeling to indicate which pair-wise comparisons display statistical significance; and 2) to graph the mean differences directly, usually with confidence intervals, and indicate in a legend or label which differences are statistically significant. The first method is a more easily interpreted choice since it resembles and depends upon typical summary statistical reporting. The second method is a more easily constructed figure (*c.f.*, Example Figure 4), but requires more interpretation and does not have as much added value over a numeric summary.

When the method of graphing group means and labeling significant differences is used, the simplest method for indicating significance of pair-wise post-hoc comparisons is to sort the means in ascending or descending order in the figure and draw “over bars” in the plot area spanning any means which are statistically equivalent (*i.e.*, not significantly different). At the further reaches of any set of equivalent means, multiple over bars may overlap. An alternative method is to indicate statistically equivalent subsets of means by attaching lettered labels to each value within a subset of equivalent means, one letter per subset. Letterings may overlap.

Example Figure 5

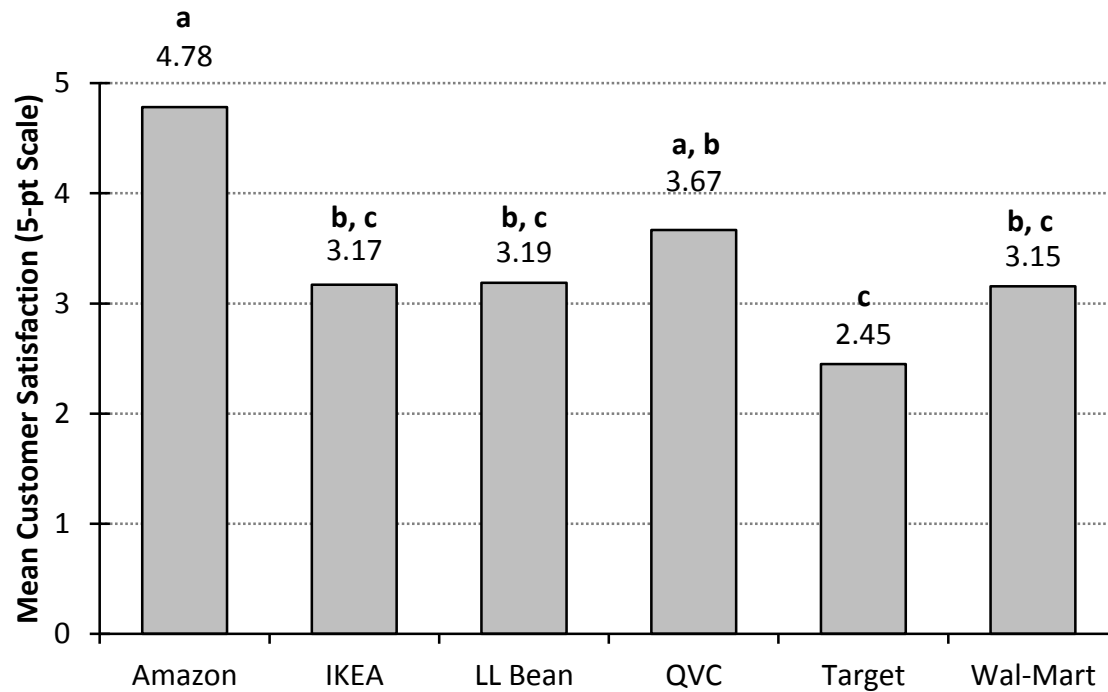
*Comparison of Mean Page Load Times (Milliseconds) for Six Major Websites*



*Note: Error bars represent 95% confidence interval of estimated mean page load times  
Over bars indicate subsets of group means that are statistically equivalent ( $p > 0.05$ )*

Example Figure 6

*Comparison of Mean Customer Satisfaction (5-pt Scale) for Six Online Retailers*



*Note: Subsets of group means that share letter labels are statistically equivalent ( $p > 0.05$ )*

## REPORTING CHI-SQUARE TESTS OF DISTRIBUTIONS

When the result of a Chi-Square goodness-of-fit test or Chi-Square test-for-independence is reported, the usual method is to provide a frequency distribution table or cross-tabulation of the observed distribution and a brief written summary of the test result. It is not required to report the expected distribution, but if it is desired to report expected frequencies the best choice is to present expected frequencies immediately below or beside observed frequencies rather than in a separate table (*c.f.*, Example Table 4 for a complete example of such a table of observed vs. expected frequencies). Chi-Square test statistics should be reported with degrees of freedom and sample size of the test indicated in parentheses attached to the  $\chi^2$  designator—*e.g.*, **based on the following reported online shopper ages and preferred retailers, a significant relationship was found between age group and preferred store,  $\chi^2 (15, N = 168) = 31.4, p < 0.01$ .**

Example Table 11

*Frequencies of Reported Preferred Online Retailer by Shopper Age Group*

Age Group	Amazon	Ikea	LL Bean	QVC	Target	Wal-Mart	Total
20 - 29 years	16	10	4	3	6	4	43
30 - 39 years	19	6	6	4	10	5	50
40 - 49 years	7	5	6	8	6	11	43
Over 50 years	4	3	4	12	4	5	32
Total	46	24	20	27	26	25	168

## REPORTING CORRELATION & LINEAR REGRESSION

The summary statistic of interest for bivariate or multivariate statistical analysis is(are) the correlation coefficient(s) between variables. When reported within the body of a report, the correlation coefficient should be presented as a decimal value and formatted appropriately for that presentation. Because correlation coefficients can never exceed a value of 1, and in reality never actually take such a value, the leading zero from the decimal value may be dropped—*e.g.*, the correlation between the tuition cost and the number of applications submitted to the 700 colleges sampled was  $-.188$ ; high tuition corresponded slightly to lower applicant pools.

When a large quantity of correlation coefficients are to be reported (*e.g.*, multivariate analysis), the correlation coefficients are best presented in a table called a correlation matrix. The column and row headers of a correlation matrix are the same and correspond to the variables for which correlation coefficients were calculated. The entries in the table are the calculated correlation coefficients for the variable pairings associated with each cell or intersection in the table. The correlation matrix is symmetric, so only one half of the matrix should be reported. In addition, the diagonal will consist only of 1s (correlation of a variable with itself) and should be omitted.

Example Table 12

*Correlation Matrix between Factors Influencing Undergraduate Student Application Decisions*

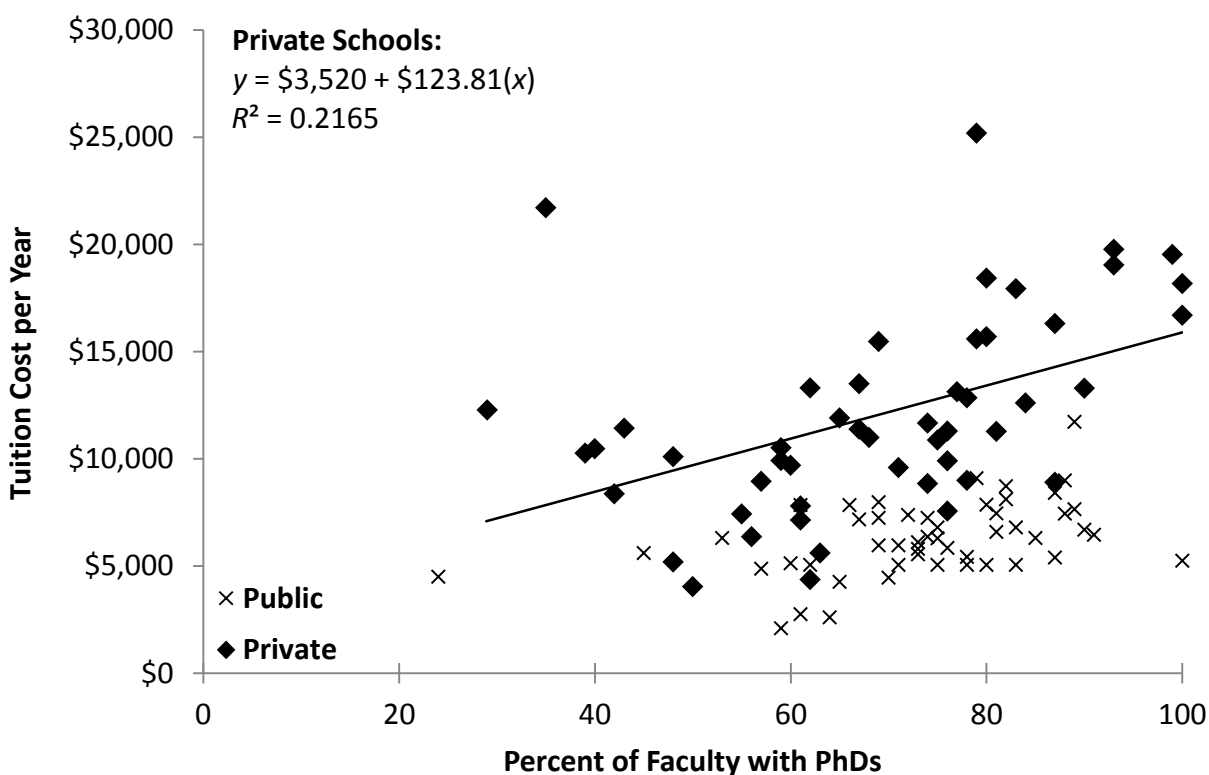
	<i>Tuition</i>	<i>SAT</i>	<i>Accept.</i>	<i>Grad.</i>	<i>% PhDs</i>	<i>Stu-Fac</i>
<i>Tuition (\$ / Year)</i>	----					
<i>Average SAT Score</i>	.62	----				
<i>% Acceptance Rate</i>	-.15	-.40	----			
<i>% Graduation Rate</i>	.59	.58	-.21	----		
<i>% Faculty with PhDs</i>	.32	.58	-.28	.26	----	
<i>Student-Faculty Ratio</i>	-.43	-.30	.06	-.29	-.15	----

If significance testing was conducted on individual correlation coefficients, it may be reported in the narrative directly for single coefficients of interest, with the degrees of freedom indicated in parentheses attached to the  $r$  designator—*e.g.*, correlation between acceptance and graduation rates was found to be inverse as expected and was significant,  $r(698) = -.21, p < 0.001$ . If multiple correlation coefficients were tested, the significance of each coefficient should be indicated in the correlation matrix with asterisks as appropriate—*e.g.*, \*  $p < 0.05$ , \*\*  $p < 0.01$ . Alternatively, a threshold absolute value of significance may be calculated and reported in a note to the table.

When the relationship between two variables of interest is reported graphically in a scatterplot, the guidelines for the general preparation of charts and figures should be observed. Additionally, care should be taken to ensure that any supposed causal relationship between the two variables is reflected in the assignment of dependent ( $Y$ ) and independent ( $X$ ) variables. If multiple series of data are presented together, the markers used for each series should be easily differentiated and a legend should be given. If a trend line is presented on the chart, the estimated regression equation and  $R^2$  value may be noted. The line should not extend beyond observed data points.

Example Figure 7

Scatterplot of Percent of Faculty with PhDs against Tuition Cost per Year, by School Classification



#### Reporting Linear Regression Estimates

When reported directly in the body of a report, the result of a linear regression estimate may be given either in terms of the effect of a single predictor (*i.e.*, using the estimated slope / beta coefficient) or in terms of the overall modeled relationship between all predictors (*i.e.*, using the estimated model *F* test statistic). In practice, it is probably best to report both. When the effect of a single predictor is reported, the estimated slope coefficient should be reported first, followed by the associated *t* test statistic and *p*-value for the coefficient—*e.g.*, the acceptance rate of a school was found to be a significant predictor of tuition cost,  $b = \$30.21$ ,  $t(696) = 3.45$ ,  $p < 0.001$ . When the entire regression model is referenced, the estimated value of *R*-squared should be reported first, followed by the associated model *F* test statistic and *p*-value—*e.g.*, the acceptance rates and average SAT scores of schools sampled were found to explain 39 percent of the variance in tuition costs with high significance,  $R^2 = .389$ ,  $F(2, 696) = 223.39$ ,  $p < 0.001$ .

When the estimated regression equation itself is to be given in the body of a report, it may be presented either abstractly using *y* and *x<sub>i</sub>* to reference the appropriate variables (as long as the variables are defined elsewhere) or in more concrete terms using the actual names or labels of the variables in the equation—*e.g.*, tuition was found to be described by the predictor variables as follows:  $Tuition = -\$13,378.48 + \$21.85 (\text{Average SAT Score}) + \$30.21 (\% \text{ Acceptance Rate})$ . Because the interpretation of estimated slope coefficients depends on the measurement units of the variables, it is critical that the units of all variables are defined somewhere in the report.



### Reporting Full Linear Regression Output

When the result of a linear regression estimate is to be reported in a table for ease of reference, there are two tables that are commonly used: 1) the ANOVA for the estimated regression model; and 2) the table of estimated coefficients and associated test statistics. The ANOVA table may be prepared following the previous guidelines for ANOVA for the comparison of group means, with only a minor labeling change to indicate the sources of variance appropriate to regression (*i.e.*, from regression and from error or residual). The table of estimated regression coefficients and test statistics should include a listing of all predictor variables included in the estimated model, with an estimated slope coefficient (standardized or unstandardized coefficients are acceptable) and either a *t* test statistic or a standard error (or both) for each estimated coefficient. The *t* test statistic is more common, and is the only acceptable choice if standardized beta coefficients are given instead of unstandardized. The significance of each coefficient should be noted clearly.

Example Table 13

*Summary of ANOVA for Regression of SAT Score and Graduation Rate against Tuition Cost*

	<i>Sum of Squares</i>	<i>df</i>	<i>Mean Square</i>	<i>F</i>
From Regression	5,197,669,758	2	2,598,834,879	296.10***
From Residuals	6,108,710,940	696	8,776,884	
Overall	11,306,380,699	698		

*Note: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ; model adjusted  $R^2 = 0.458$*

Example Table 14

*Estimated Result for Regression of SAT Score and Graduation Rate against Tuition Cost*

<i>Variable</i>	<i>Coefficient</i>	<i>Std. Error</i>	<i>t Statistic</i>	<i>95% Confidence Interval</i>	
				<i>Lower Limit</i>	<i>Upper Limit</i>
Intercept	-\$7,974.64	\$918.61		-\$9,778.22	-\$6,171.06
SAT Score	\$13.79	\$1.13	12.23***	\$11.58	\$16.00
Graduation Rate	\$75.03	\$7.43	10.10***	\$60.44	\$89.61

*Note: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ; model adjusted  $R^2 = 0.458$*

## REPORTING MULTIPLE COMPETING REGRESSION MODELS

When multiple regression models are compared according to the associated model *F* statistics for significance and  $R^2$  fit statistics, either as a result of multiple samples to be analyzed using the same model or as a result of multiple model specifications to be tested against each other, the results of the competing regressions should be reported in a single table. The table should list all variables used as predictors in any regression model specification, and should include a column for estimated coefficients and test statistics under each model estimate (with a label spanning each model's associated columns). Model  $R^2$  or *F* values are reported in the final row.

Example Table 15

Comparison of Estimated Results for Model 2 in Private vs. Public Schools

Variable	Private Schools (n = 476)		Public Schools (n = 222)	
	Coef.	t Stat	Coef.	t Stat
Intercept	-\$7,968.67		-\$1,406.28	
SAT Score	\$18.58	15.56***	\$4.62	2.85**
Graduation Rate	\$26.79	3.32***	\$52.40	5.01***
Acceptance Rate	-\$6.08	-0.67	\$15.51	1.70
Model R <sup>2</sup>   F stat	0.52	175.44***	0.26	26.93***

Note: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Example Table 16

Comparison of Estimated Results for Model 1, 2, and 3 Predicting School Tuition Costs

Variable	Model 1		Model 2		Model 3	
	Coef.	t Stat	Coef.	t Stat	Coef.	t Stat
Intercept	-\$7,974.64		-\$11,380.55		-\$6,306.15	
SAT Score	\$13.79	12.23***	\$15.21	12.74***	\$13.68	10.13***
Graduation Rate	\$75.03	10.10***	\$74.34	10.08***	\$65.93	9.17***
Acceptance Rate			\$27.95	3.41***	\$23.47	2.98**
Faculty with PhDs					-\$2.18	-0.28
Student-Faculty					-\$176.74	-7.88***
Model R <sup>2</sup>   F stat	0.458	296.10***	0.466	204.31***	0.509	145.70***

Note: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

When a very large number of model specifications are to be compared, an alternative method of presentation is to give the  $t$  test statistic or standard error for each estimated coefficient in parentheses below the reported coefficient (c.f., Example Table 3) and indicate this in the note.

