ECO220Y1Y, Test #5, Prof. Murdock

April 5, 2019, 9:10 – 11:00 am

U of T E-MAIL: _____@MAIL.UTORONTO.CA

SURNAME (LAST NAME):																		
GIVEN NAME (FIRST NAME):																		

UTORID: (e.g. LIHAO118)							

Instructions:

- You have 110 minutes. Keep these test papers and the *Supplement* closed and face up on your desk until the start of the test is announced. You must stay for a minimum of 60 minutes.
- You may use a non-programmable calculator.
- There are 5 questions (most with multiple parts) with varying point values worth a total of 100 points.
- This test includes these 8 pages plus the *Supplement*. The *Supplement* contains the aid sheets (formulas, Normal, Student t, and F tables) and readings, figures, tables, and other materials required for some test questions. For each question referencing this *Supplement*, carefully review *all* materials. ***The Supplement will NOT be graded:*** write your answers on these test papers. When we announce the end of the test, hand these test papers to us (you keep the *Supplement*).
- Write your answers clearly, completely and concisely in the designated space provided immediately after each question. An answer guide ends each question to let you know what is expected. For example, a quantitative analysis (which shows your work), a fully-labelled graph, and/or sentences.
    - Anything requested by the question and/or the answer guide is required. Similarly, limit yourself to the answer guide. For example, if the answer guide does not request sentences, provide only what is requested (e.g. quantitative analysis).
    - Marking TAs are instructed to accept all reasonable rounding.
- ***Your entire answer must fit in the designated space provided immediately after each question.*** No extra space/pages are possible. You *cannot* use blank space for other questions nor can you write answers on the *Supplement*. ***Write in PENCIL and use an ERASER as needed*** so that you can fit your final answer (including work and reasoning) in the appropriate space. Questions give more blank space than is needed for an answer (with typical handwriting) worth full marks. ***Follow the answer guides and avoid excessively long answers.***

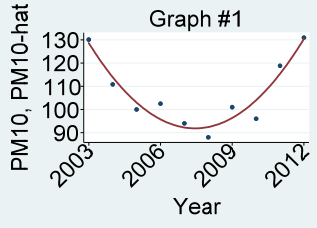
(1) See ***Supplement for Question (1)***: *The 2018 World Happiness Report.*

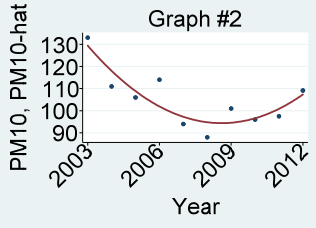
(a) [10 pts] See **Figure 2.2** and, below it, the details for Canada and Japan. What is the 95% CI estimate of the *DIFFERENCE* in mean happiness between Canada and Japan? Answer with a quantitative analysis.

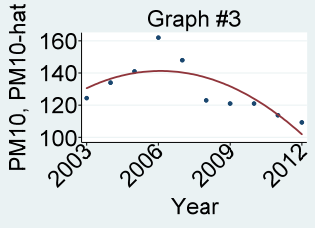
(b) [6 pts] See **Table A7**. Define $abroad$ to equal 1 for respondents with family abroad and 0 otherwise. Define $happiness$ as the life evaluation score (0-10 scale). What would be the OLS equation: $\widehat{happiness} = b_0 + b_1 abroad$? Also, what would be the sample size (n) for that OLS regression? Answer with the values of b_0, b_1, and n.

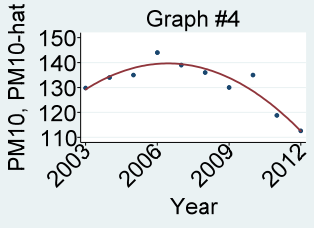
(2) See ***Supplement for Question (2)***: *Air Pollution in Tianjin, China*.

(a) [6 pts] Which of these summarizes the regression results for Tianjin? *Explain*. Answer with 2 – 3 sentences.









(b) [4 pts] How should we interpret the value of -14.73599? Answer with 1 – 2 sentences.

(3) [18 pts] See ***Supplement for Question (3)***: *California Energy*. Compare and contrast the results in boldface for **constr_01_04** in Regression #1 and Regression #2. For each, *interpret* the results in boldface. Also, explain *why* the results in boldface are similar or different across the regressions. Which one of these two regressions offers a better answer to the primary research question in this journal article? *Why* is it better? Answer with 8 – 10 sentences.

(4) See ***Supplement for Question (4)***: *Correlation matrix*.

(a) [5 pts] Is the correlation between y and x1 statistically significant? If so, at which of these common significance levels: 10%, 5%, 1%, or 0.1%? Answer with a quantitative analysis & 1 sentence.

(b) [5 pts] Is the correlation between y and x2 statistically significant? If so, at which of these common significance levels: 10%, 5%, 1%, or 0.1%? Answer with a quantitative analysis & 1 sentence.

(c) [7 pts] **True/False/Explain:** "A multiple regression of y on x1, x2, and x3 would allow us to check how y is related to each x variable and whether or not each correlation is statistically significant." Answer with 2 – 3 sentences.

(5) See ***Supplement for Question (5)***: *Parents' Beliefs About Their Children's Academic Ability*.

(a) [14 pts] Is there a statistically significant difference between the control group and treatment group in the mean **overall score**? What is the P-value? Is the answer about whether or not there is a statistically significant difference surprising or expected? *Explain*. Answer with hypotheses in formal notation, a quantitative analysis & 2 – 3 sentences.

(b) [8 pts] Using Regression (1) in **Table 1**, draw ONE graph showing how predicted endline beliefs relate to overall scores for each group: the control group and treatment group. Label the axes, specify which line belongs to which group, and clearly write the numeric values of the intercept and slope of each. Answer with a fully-labelled graph.

(c) [8 pts] What is the *model* for Regression (1) in **Table 1**? Next, continuing with Part **(b)**, is there a statistically significant difference in the slopes of the two lines? Answer with a formal regression model, hypotheses in formal notation, a quantitative analysis & 1 sentence.

(d) [9 pts] All things considered, which column of results in **Table 1** is the one that the reader should focus on? *Why*? Make sure to include consideration of the R^2 in your assessment. Answer with a clear choice of Column (1), (2), or (3) & 3 – 4 sentences.

This *Supplement* contains the aid sheets (formulas, Normal, Student t, and F tables) and readings, figures, tables, and other materials for some test questions. For each question referencing this *Supplement*, carefully review *all* materials.

Sample mean: $\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$ **Sample variance:** $s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{X})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2}{n-1} - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}$ **Sample s.d.:** $s = \sqrt{s^2}$

Sample coefficient of variation: $CV = \frac{s}{\bar{X}}$ **Sample covariance:** $s_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{X})(y_i-\bar{Y})}{n-1} = \frac{\sum_{i=1}^{n} x_i y_i}{n-1} - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n(n-1)}$

Sample interquartile range: $IQR = Q3 - Q1$ **Sample coefficient of correlation:** $r = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^{n} z_{x_i} z_{y_i}}{n-1}$

Addition rule: $P(A\ or\ B) = P(A) + P(B) - P(A\ and\ B)$ **Conditional probability:** $P(A|B) = \frac{P(A\ and\ B)}{P(B)}$

Complement rules: $P(A^C) = P(A') = 1 - P(A)$ $P(A^C|B) = P(A'|B) = 1 - P(A|B)$

Multiplication rule: $P(A\ and\ B) = P(A|B)P(B) = P(B|A)P(A)$

Expected value: $E[X] = \mu = \sum_{all\ x} xp(x)$ **Variance:** $V[X] = E[(X-\mu)^2] = \sigma^2 = \sum_{all\ x}(x-\mu)^2 p(x)$

Covariance: $COV[X,Y] = E[(X-\mu_X)(Y-\mu_Y)] = \sigma_{XY} = \sum_{all\ x}\sum_{all\ y}(x-\mu_X)(y-\mu_Y)p(x,y)$

Laws of expected value:

$E[c] = c$

$E[X + c] = E[X] + c$

$E[cX] = cE[X]$

$E[a + bX + cY] = a + bE[X] + cE[Y]$

Laws of variance:

$V[c] = 0$

$V[X + c] = V[X]$

$V[cX] = c^2 V[X]$

$V[a + bX + cY] = b^2 V[X] + c^2 V[Y] + 2bc * COV[X,Y]$

$V[a + bX + cY] = b^2 V[X] + c^2 V[Y] + 2bc * SD(X) * SD(Y) * \rho$

where $\rho = CORRELATION[X,Y]$

Laws of covariance:

$COV[X, c] = 0$

$COV[a + bX, c + dY] = bd * COV[X,Y]$

Combinatorial formula: $C_x^n = \frac{n!}{x!(n-x)!}$ **Binomial probability:** $p(x) = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$ for $x = 0,1,2,\ldots,n$

If *X* is Binomial $(X \sim B(n,p))$ **then** $E[X] = np$ **and** $V[X] = np(1-p)$

If *X* is Uniform $(X \sim U[a,b])$ **then** $f(x) = \frac{1}{b-a}$ **and** $E[X] = \frac{a+b}{2}$ **and** $V[X] = \frac{(b-a)^2}{12}$

Sampling distribution of $\bar{X}$:

$\mu_{\bar{X}} = E[\bar{X}] = \mu$

$\sigma_{\bar{X}}^2 = V[\bar{X}] = \frac{\sigma^2}{n}$

$\sigma_{\bar{X}} = SD[\bar{X}] = \frac{\sigma}{\sqrt{n}}$

Sampling distribution of $\hat{P}$:

$\mu_{\hat{P}} = E[\hat{P}] = p$

$\sigma_{\hat{P}}^2 = V[\hat{P}] = \frac{p(1-p)}{n}$

$\sigma_{\hat{P}} = SD[\hat{P}] = \sqrt{\frac{p(1-p)}{n}}$

Sampling distribution of $(\hat{P}_2 - \hat{P}_1)$:

$\mu_{\hat{P}_2-\hat{P}_1} = E[\hat{P}_2 - \hat{P}_1] = p_2 - p_1$

$\sigma_{\hat{P}_2-\hat{P}_1}^2 = V[\hat{P}_2 - \hat{P}_1] = \frac{p_2(1-p_2)}{n_2} + \frac{p_1(1-p_1)}{n_1}$

$\sigma_{\hat{P}_2-\hat{P}_1} = SD[\hat{P}_2 - \hat{P}_1] = \sqrt{\frac{p_2(1-p_2)}{n_2} + \frac{p_1(1-p_1)}{n_1}}$

Sampling distribution of $(\bar{X}_1 - \bar{X}_2)$, independent samples:

$\mu_{\bar{X}_1-\bar{X}_2} = E[\bar{X}_1 - \bar{X}_2] = \mu_1 - \mu_2$

$\sigma_{\bar{X}_1-\bar{X}_2}^2 = V[\bar{X}_1 - \bar{X}_2] = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

$\sigma_{\bar{X}_1-\bar{X}_2} = SD[\bar{X}_1 - \bar{X}_2] = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Sampling distribution of $(\bar{X}_d)$, paired $(d = X_1 - X_2)$:

$\mu_{\bar{X}_d} = E[\bar{X}_d] = \mu_1 - \mu_2$

$\sigma_{\bar{X}_d}^2 = V[\bar{X}_d] = \frac{\sigma_d^2}{n} = \frac{\sigma_1^2 + \sigma_2^2 - 2*\rho*\sigma_1*\sigma_2}{n}$

$\sigma_{\bar{X}_d} = SD[\bar{X}_d] = \frac{\sigma_d}{\sqrt{n}} = \sqrt{\frac{\sigma_1^2 + \sigma_2^2 - 2*\rho*\sigma_1*\sigma_2}{n}}$

Inference about a population proportion:

***z* test statistic:** $z = \frac{\hat{P}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ **CI estimator:** $\hat{P} \pm z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$

Inference about comparing two population proportions:

***z* test statistic under Null hypothesis of no difference:** $z = \frac{\hat{P}_2-\hat{P}_1}{\sqrt{\frac{\bar{P}(1-\bar{P})}{n_1}+\frac{\bar{P}(1-\bar{P})}{n_2}}}$ **Pooled proportion:** $\bar{P} = \frac{X_1+X_2}{n_1+n_2}$

CI estimator: $(\hat{P}_2 - \hat{P}_1) \pm z_{\alpha/2}\sqrt{\frac{\hat{P}_2(1-\hat{P}_2)}{n_2} + \frac{\hat{P}_1(1-\hat{P}_1)}{n_1}}$

Inference about the population mean:

***t* test statistic:** $t = \frac{\bar{X}-\mu_0}{s/\sqrt{n}}$ **CI estimator:** $\bar{X} \pm t_{\alpha/2}\,\frac{s}{\sqrt{n}}$ **Degrees of freedom:** $\nu = n - 1$

Inference about a comparing two population means, independent samples, unequal variances:

***t* test statistic:** $t = \frac{(\bar{X}_1-\bar{X}_2)-\Delta_0}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$ **CI estimator:** $(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Degrees of freedom: $\nu = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{s_1^2}{n_1}\right)^2+\frac{1}{n_2-1}\left(\frac{s_2^2}{n_2}\right)^2}$

Inference about a comparing two population means, independent samples, assuming equal variances:

***t* test statistic:** $t = \frac{(\bar{X}_1-\bar{X}_2)-\Delta_0}{\sqrt{\frac{s_p^2}{n_1}+\frac{s_p^2}{n_2}}}$ **CI estimator:** $(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2}\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$ **Degrees of freedom:** $\nu = n_1 + n_2 - 2$

Pooled variance: $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$

Inference about a comparing two population means, paired data: (n is number of pairs and $d = X_1 - X_2$)

***t* test statistic:** $t = \frac{\bar{d}-\Delta_0}{s_d/\sqrt{n}}$ **CI estimator:** $\bar{X}_d \pm t_{\alpha/2}\,\frac{s_d}{\sqrt{n}}$ **Degrees of freedom:** $\nu = n - 1$

SIMPLE REGRESSION:

Model: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ **OLS line:** $\hat{y}_i = b_0 + b_1 x_i$ $b_1 = \frac{s_{xy}}{s_x^2} = r\frac{s_y}{s_x}$ $b_0 = \bar{Y} - b_1\bar{X}$

Coefficient of determination: $R^2 = (r)^2$ **Residuals:** $e_i = y_i - \hat{y}_i$

Standard deviation of residuals: $s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{\sum_{i=1}^{n}(e_i-0)^2}{n-2}}$ **Standard error of slope:** $s.e.(b_1) = s_{b_1} = \frac{s_e}{\sqrt{(n-1)s_x^2}}$

Inference about the population slope:

***t* test statistic:** $t = \frac{b_1 - \beta_{10}}{s.e.(b_1)}$ **CI estimator:** $b_1 \pm t_{\alpha/2} s.e.(b_1)$ **Degrees of freedom:** $\nu = n - 2$

Standard error of slope: $s.e.(b_1) = s_{b_1} = \frac{s_e}{\sqrt{(n-1)s_x^2}}$

Prediction interval for y at given value of x (x_g):

$\hat{y}_{x_g} \pm t_{\alpha/2} s_e \sqrt{1 + \frac{1}{n} + \frac{(x_g - \bar{X})^2}{(n-1)s_x^2}}$ **or** $\hat{y}_{x_g} \pm t_{\alpha/2} \sqrt{(s.e.(b_1))^2 (x_g - \bar{X})^2 + \frac{s_e^2}{n} + s_e^2}$

Degrees of freedom: $\nu = n - 2$

Confidence interval for predicted mean at given value of x (x_g):

$\hat{y}_{x_g} \pm t_{\alpha/2} s_e \sqrt{\frac{1}{n} + \frac{(x_g - \bar{X})^2}{(n-1)s_x^2}}$ **or** $\hat{y}_{x_g} \pm t_{\alpha/2} \sqrt{(s.e.(b_1))^2 (x_g - \bar{X})^2 + \frac{s_e^2}{n}}$ **Degrees of freedom:** $\nu = n - 2$

SIMPLE & MULTIPLE REGRESSION:

Model: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i$

$SST = \sum_{i=1}^{n} (y_i - \bar{Y})^2 = SSR + SSE \quad SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{Y})^2 \quad SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

$s_y^2 = \frac{SST}{n-1} \quad MSE = \frac{SSE}{n-k-1} \quad Root\ MSE = \sqrt{\frac{SSE}{n-k-1}} \quad MSR = \frac{SSR}{k}$

$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad Adj.R^2 = 1 - \frac{SSE/(n-k-1)}{SST/(n-1)} = \left(R^2 - \frac{k}{n-1}\right)\left(\frac{n-1}{n-k-1}\right)$

Residuals: $e_i = y_i - \hat{y}_i$ **Standard deviation of residuals:** $s_e = \sqrt{\frac{SSE}{n-k-1}} = \sqrt{\frac{\sum_{i=1}^{n}(e_i - 0)^2}{n-k-1}}$

Inference about the overall statistical significance of the regression model:

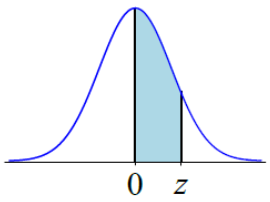
$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} = \frac{(SST-SSE)/k}{SSE/(n-k-1)} = \frac{SSR/k}{SSE/(n-k-1)} = \frac{MSR}{MSE}$

Numerator degrees of freedom: $\nu_1 = k$ **Denominator degrees of freedom:** $\nu_2 = n - k - 1$

Inference about the population slope for explanatory variable j:

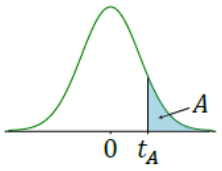
***t* test statistic:** $t = \frac{b_j - \beta_{j0}}{s_{b_j}}$ **CI estimator:** $b_j \pm t_{\alpha/2} s_{b_j}$ **Degrees of freedom:** $\nu = n - k - 1$

Standard error of slope: $s.e.(b_j) = s_{b_j}$ (for multiple regression, must be obtained from technology)



The Standard Normal Distribution:

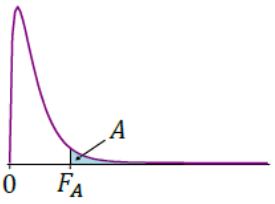
	Second decimal place in z									
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999



Critical Values of Student t Distribution:

ν	$t_{0.10}$	$t_{0.05}$	$t_{0.025}$	$t_{0.01}$	$t_{0.005}$	$t_{0.001}$	$t_{0.0005}$	ν	$t_{0.10}$	$t_{0.05}$	$t_{0.025}$	$t_{0.01}$	$t_{0.005}$	$t_{0.001}$	$t_{0.0005}$
1	3.078	6.314	12.71	31.82	63.66	318.3	636.6	38	1.304	1.686	2.024	2.429	2.712	3.319	3.566
2	1.886	2.920	4.303	6.965	9.925	22.33	31.60	39	1.304	1.685	2.023	2.426	2.708	3.313	3.558
3	1.638	2.353	3.182	4.541	5.841	10.21	12.92	40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610	41	1.303	1.683	2.020	2.421	2.701	3.301	3.544
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869	42	1.302	1.682	2.018	2.418	2.698	3.296	3.538
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959	43	1.302	1.681	2.017	2.416	2.695	3.291	3.532
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408	44	1.301	1.680	2.015	2.414	2.692	3.286	3.526
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041	45	1.301	1.679	2.014	2.412	2.690	3.281	3.520
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781	46	1.300	1.679	2.013	2.410	2.687	3.277	3.515
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587	47	1.300	1.678	2.012	2.408	2.685	3.273	3.510
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437	48	1.299	1.677	2.011	2.407	2.682	3.269	3.505
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318	49	1.299	1.677	2.010	2.405	2.680	3.265	3.500
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221	50	1.299	1.676	2.009	2.403	2.678	3.261	3.496
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140	51	1.298	1.675	2.008	2.402	2.676	3.258	3.492
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073	52	1.298	1.675	2.007	2.400	2.674	3.255	3.488
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015	53	1.298	1.674	2.006	2.399	2.672	3.251	3.484
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965	54	1.297	1.674	2.005	2.397	2.670	3.248	3.480
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922	55	1.297	1.673	2.004	2.396	2.668	3.245	3.476
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883	60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850	65	1.295	1.669	1.997	2.385	2.654	3.220	3.447
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819	70	1.294	1.667	1.994	2.381	2.648	3.211	3.435
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792	75	1.293	1.665	1.992	2.377	2.643	3.202	3.425
23	1.319	1.714	2.069	2.500	2.807	3.485	3.768	80	1.292	1.664	1.990	2.374	2.639	3.195	3.416
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745	90	1.291	1.662	1.987	2.368	2.632	3.183	3.402
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725	100	1.290	1.660	1.984	2.364	2.626	3.174	3.390
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707	120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690	140	1.288	1.656	1.977	2.353	2.611	3.149	3.361
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674	160	1.287	1.654	1.975	2.350	2.607	3.142	3.352
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659	180	1.286	1.653	1.973	2.347	2.603	3.136	3.345
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646	200	1.286	1.653	1.972	2.345	2.601	3.131	3.340
31	1.309	1.696	2.040	2.453	2.744	3.375	3.633	250	1.285	1.651	1.969	2.341	2.596	3.123	3.330
32	1.309	1.694	2.037	2.449	2.738	3.365	3.622	300	1.284	1.650	1.968	2.339	2.592	3.118	3.323
33	1.308	1.692	2.035	2.445	2.733	3.356	3.611	400	1.284	1.649	1.966	2.336	2.588	3.111	3.315
34	1.307	1.691	2.032	2.441	2.728	3.348	3.601	500	1.283	1.648	1.965	2.334	2.586	3.107	3.310
35	1.306	1.690	2.030	2.438	2.724	3.340	3.591	750	1.283	1.647	1.963	2.331	2.582	3.101	3.304
36	1.306	1.688	2.028	2.434	2.719	3.333	3.582	1000	1.282	1.646	1.962	2.330	2.581	3.098	3.300
37	1.305	1.687	2.026	2.431	2.715	3.326	3.574	∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

Degrees of freedom: ν



The F Distribution:

ν_1	1	2	3	4	5	6	7	8	9	10	11	12	15	20	30	∞
ν_2	**Critical Values of F Distribution for $A = 0.10$:**															
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.28	3.27	3.24	3.21	3.17	3.10
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.30	2.28	2.24	2.20	2.16	2.06
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.04	2.02	1.97	1.92	1.87	1.76
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.91	1.89	1.84	1.79	1.74	1.61
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.79	1.77	1.72	1.67	1.61	1.46
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.74	1.71	1.66	1.61	1.54	1.38
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.68	1.66	1.60	1.54	1.48	1.29
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.63	1.60	1.55	1.48	1.41	1.19
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.57	1.55	1.49	1.42	1.34	1.00
ν_2	**Critical Values of F Distribution for $A = 0.05$:**															
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.62	4.56	4.50	4.36
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.85	2.77	2.70	2.54
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48	2.40	2.33	2.25	2.07
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28	2.20	2.12	2.04	1.84
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09	2.01	1.93	1.84	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00	1.92	1.84	1.74	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.84	1.75	1.65	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.87	1.83	1.75	1.66	1.55	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.79	1.75	1.67	1.57	1.46	1.00
ν_2	**Critical Values of F Distribution for $A = 0.01$:**															
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.96	9.89	9.72	9.55	9.38	9.02
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.77	4.71	4.56	4.41	4.25	3.91
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67	3.52	3.37	3.21	2.87
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.29	3.23	3.09	2.94	2.78	2.42
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84	2.70	2.55	2.39	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66	2.52	2.37	2.20	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50	2.35	2.20	2.03	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.40	2.34	2.19	2.03	1.86	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.25	2.18	2.04	1.88	1.70	1.00
ν_2	**Critical Values of F Distribution for $A = 0.001$:**															
5	47.2	37.1	33.2	31.1	29.8	28.8	28.2	27.6	27.2	26.9	26.6	26.4	25.9	25.4	24.9	23.8
10	21.0	14.9	12.6	11.3	10.5	9.93	9.52	9.20	8.96	8.75	8.59	8.45	8.13	7.80	7.47	6.76
15	16.6	11.3	9.34	8.25	7.57	7.09	6.74	6.47	6.26	6.08	5.94	5.81	5.54	5.25	4.95	4.31
20	14.8	9.95	8.10	7.10	6.46	6.02	5.69	5.44	5.24	5.08	4.94	4.82	4.56	4.29	4.00	3.38
30	13.3	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.39	4.24	4.11	4.00	3.75	3.49	3.22	2.59
40	12.6	8.25	6.59	5.70	5.13	4.73	4.44	4.21	4.02	3.87	3.75	3.64	3.40	3.14	2.87	2.23
60	12.0	7.77	6.17	5.31	4.76	4.37	4.09	3.86	3.69	3.54	3.42	3.32	3.08	2.83	2.55	1.89
120	11.4	7.32	5.78	4.95	4.42	4.04	3.77	3.55	3.38	3.24	3.12	3.02	2.78	2.53	2.26	1.54
∞	10.83	6.91	5.42	4.62	4.10	3.74	3.47	3.27	3.10	2.96	2.84	2.74	2.51	2.27	1.99	1.00

Numerator degrees of freedom: ν_1; Denominator degrees of freedom: ν_2

Supplement for Question (1): The *2018 World Happiness Report* by John F. Helliwell, Richard Layard, and Jeffrey D. Sachs (http://worldhappiness.report/) documents and analyzes individual well-being across countries and over time.

> **Excerpt (p. 104):** [We measure happiness, also called "life evaluations," with] the question "Please imagine a ladder with steps numbered from zero at the bottom to ten at the top. Suppose we say that the top of the ladder represents the best possible life for you, and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time, assuming that the higher the step the better you feel about your life, and the lower the step the worse you feel about it? Which step comes closest to the way you feel?"

Surveys ask this Cantril ladder question to a fresh random sample of people in each year and each country. Below are pieces of Figure 2.2 and an excerpt that explains it. (The full figure is three pages long and shows 156 countries.)

> **Excerpt (p. 21):** Figure 2.2 shows the average ladder score (answer to the Cantril ladder question, on a scale of 0 to 10) for each country, averaged over the years 2015-2017. The total sample sizes are reported in the statistical appendix, and are reflected in Figure 2.2 by the horizontal lines showing the 95% confidence regions.

Figure 2.2: Ranking of Happiness 2015–2017 (Part 1)

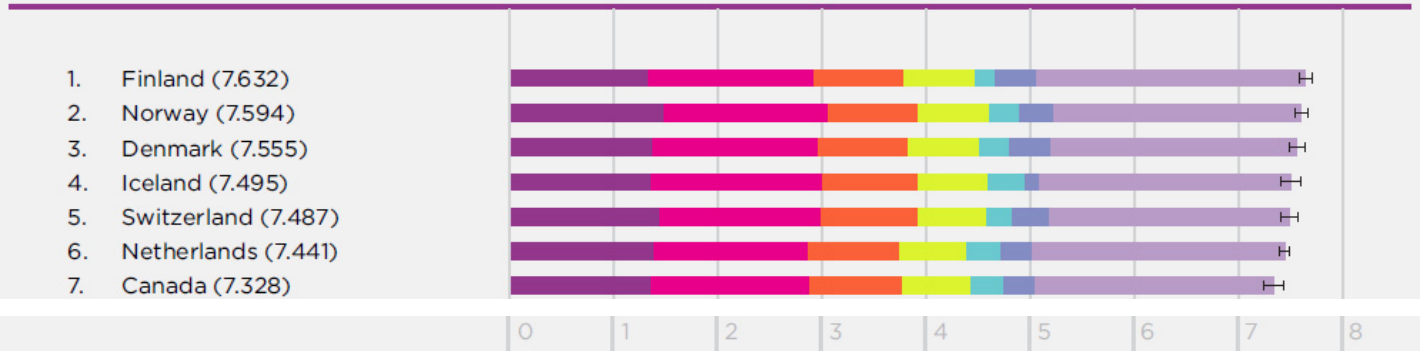
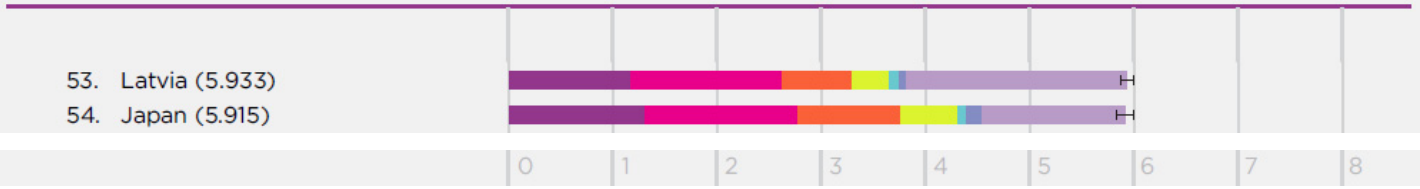


Figure 2.2: Ranking of Happiness 2015–2017 (Part 2)



Details for Canada and Japan: For 2015-2017 combined, the sample size for Canada (ranked 7) is 2,027 and the values of the 95% CI shown in the figure are: [7.2363, 7.4207]. The sample size for Japan (ranked 54) is 3,008 and the values of the 95% CI shown in the figure are: [5.8333, 5.9967].

Below is an excerpt of Table A7 on p. 109.

Table A7: Summary Statistics for Respondents with and Without Relative Abroad

	No family abroad, N=19,933		Family abroad, N=3,976	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
Live evaluations (0-10 scale)	6.414	2.305	6.336	2.287

Notes: Includes Venezuela, Brazil, Mexico, Costa Rica, Argentina, Bolivia, Chile, Colombia, Ecuador, El Salvador, Guatemala, Honduras, Nicaragua, Panama, Paraguay, Peru, and Uruguay and excludes the foreign-born in each country of interview.

Supplement for Question (2): Recall Zheng and Kahn (2017) "A New Era of Pollution Progress in Urban China?" PM10 measures of air pollution in micrograms per cubic meter of air ($\mu g/m^3$). The regression below uses 10 years of data, 2003 through 2012, for the city of Tianjin. The variable `trend` is 1 for the year 2003, 2 for 2004, 3 for 2005, …, and 10 for 2012. The variable `trend_sq` is trend squared. One value below is in boldface for easy reference.

```
      Source |       SS           df       MS      Number of obs   =        10
-------------+----------------------------------   F(2, 7)         =     12.29
       Model |  1153.57945         2  576.789726   Prob > F        =    0.0051
    Residual |  328.492494         7  46.9274991   R-squared       =    0.7784
-------------+----------------------------------   Adj R-squared   =    0.7150
       Total |  1482.07195         9  164.674661   Root MSE        =    6.8504

------------------------------------------------------------------------------
        pm10 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       trend |  -14.73599   3.364972    -4.38   0.003    -22.69289   -6.779099
    trend_sq |       1.117   .2981239     3.75   0.007      .412049    1.821951
       _cons |   143.0097    8.05707    17.75   0.000     123.9577    162.0616
------------------------------------------------------------------------------
```

Supplement for Question (3): Recall Levinson (2016) "How Much Energy Do Building Energy Codes Save? Evidence from California Houses" (https://www.aeaweb.org/articles?id=10.1257/aer.20150102). Below are two multiple regressions from Excel with parts in boldface for easy reference. In both Regression #1 and #2, the y variable is ln_elec_mmbtu, which is the natural log of annual household electricity use in MMBTUs. yr_2009 is a survey year dummy (=1 if 2009 RASS data, =0 if 2003 RASS data). A set of dummy variables record when the house was constructed, with before 1940 serving as the reference (omitted) category. (For example, constr_40_49 =1 if constructed from 1940-1949, =0 otherwise.)

Regression #1: (Dependent variable is the natural logarithm of annual household electricity use in MMBTUs)

Regression Statistics	
R Squared	0.086038686
Observations	14045

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	12	316.1656134	26.34713445	110.0789512	4.9946E-263
Residual	14032	3358.525736	0.239347615		
Total	14044	3674.69135			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	2.80091723	0.014998686	186.7441713	0	2.77151781	2.830316651
yr_2009	0.099643802	0.008411839	11.84566212	3.22698E-32	0.083155478	0.116132125
constr_40_49	0.026738028	0.021701186	1.232099856	0.217932431	-0.015799184	0.06927524
constr_50_59	0.110574793	0.01757221	6.292594734	3.21478E-10	0.076130924	0.145018662
constr_60_69	0.244376666	0.017853879	13.68759493	2.25322E-42	0.209380687	0.279372644
constr_70_74	0.276537056	0.020694214	13.36301306	1.75802E-40	0.235973642	0.317100469
constr_75_77	0.346758564	0.023614619	14.68406364	1.8657E-48	0.300470769	0.393046359
constr_78_82	0.366688048	0.021044291	17.42458582	2.73986E-67	0.325438439	0.407937658
constr_83_92	0.391032426	0.017953747	21.77998972	1.8423E-103	0.355840693	0.426224159
constr_93_97	0.378969248	0.022834935	16.59602942	2.85445E-61	0.334209738	0.423728758
constr_98_00	0.380513957	0.024034536	15.83196574	5.71125E-56	0.333403067	0.427624846
constr_01_04	**0.394942885**	0.025326069	15.59432225	**2.27406E-54**	0.345300419	0.44458535
constr_05_08	0.377592192	0.032209664	11.72294738	1.36488E-31	0.314456965	0.440727418

Supplement for Question (3), continues on the next page >>>>>

Supplement for Question (3), cont'd:

Regression #2 also includes the following variables: cool_deg_days (cooling degree days in 100s for that year and house location), ln_sq_feet (natural log of house size in 1,000s of square feet), ln_num_res (natural log of the number of residents living at this address), and central_ac (=1 if house has central air conditioning, =0 otherwise).

Regression #2: (Dependent variable is the natural logarithm of annual household electricity use in MMBTUs)

Regression Statistics	
R Squared	0.34926507
Observations	14045

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	16	1283.441333	80.21508331	470.5727886	0
Residual	14028	2391.250017	0.170462647		
Total	14044	3674.69135			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	2.299491392	0.014919424	154.1273549	0	2.270247335	2.32873545
cool_deg_days	0.023994715	0.001136583	21.11128134	2.03848E-97	0.021766861	0.026222568
ln_sq_feet	0.43683163	0.009618505	45.41575184	0	0.41797808	0.45568518
ln_num_res	0.256371747	0.006796996	37.71838984	1.2487E-296	0.24304873	0.269694764
central_ac	0.211006685	0.008275383	25.4981183	3.1576E-140	0.194785834	0.227227537
yr_2009	0.051487662	0.007339257	7.015378481	2.39835E-12	0.037101743	0.065873582
constr_40_49	0.029368085	0.018332682	1.601952457	0.109188639	-0.006566412	0.065302581
constr_50_59	0.037917498	0.014877628	2.548625141	0.010825406	0.008755366	0.06707963
constr_60_69	0.078540997	0.015253059	5.149196341	2.65146E-07	0.04864297	0.108439023
constr_70_74	0.077211207	0.017697734	4.362773726	1.29341E-05	0.042521293	0.11190112
constr_75_77	0.093859498	0.020283551	4.627370234	3.73654E-06	0.054101039	0.133617957
constr_78_82	0.088566174	0.018226966	4.859073978	1.19211E-06	0.052838896	0.124293453
constr_83_92	0.059345294	0.015944254	3.722048946	0.000198385	0.028092434	0.090598154
constr_93_97	0.000500876	0.020048108	0.02498371	0.980068313	-0.038796084	0.039797836
constr_98_00	-0.030807821	0.021152757	-1.456444687	0.145292111	-0.072270041	0.010654399
constr_01_04	**-0.052840257**	0.022329139	-2.366426071	**0.017974268**	-0.096608342	-0.009072172
constr_05_08	-0.131700471	0.028135997	-4.680853193	2.88357E-06	-0.18685077	-0.076550172

Supplement for Question (4): Consider the following correlation matrix for observational data with 100 observations and four variables.

```
correlate y x1 x2 x3;

(obs=100)
             |        y       x1       x2       x3
-------------+------------------------------------
           y |   1.0000
          x1 |  -0.2892   1.0000
          x2 |   0.1891  -0.0182   1.0000
          x3 |   0.1423  -0.0384   0.3216   1.0000
```

Supplement for Question (5): In the article "Parents' Beliefs About Their Children's Academic Ability: Implications for Educational Investments," (https://www.aeaweb.org/articles?id=10.1257/aer.20171172) Dizon-Ross (2019) shows how providing parents with clear, understandable information about their child's academic progress can help parents make more informed decisions. She does a field experiment with 5,268 children from 39 randomly selected primary schools in two districts (Machinga and Balaka) in Malawi. Here is a quick summary.

- The researchers have the scores (out of 100 points) for every student in the subjects of math, English, and Chichewa. The **overall score**, which measures academic performance, is the average of these three scores.
- Schools are required to send school report cards home to parents. However, these are often lost (students do not deliver them to their parents) or are unclear to uneducated parents.
- In the **baseline** survey, researchers asked parents about *beliefs* about their children's academic performance.
- The 5,268 children are randomly divided into a **treatment group** ($n_T = 2{,}614$) and **control group** ($n_C = 2{,}654$).
    - For the treatment group, parents received an additional and specially designed report card on the child's performance, which the researchers carefully designed to be clear for all parents including uneducated ones. Further, a trained person walked parents through each number.
    - For the control group, parents only had the usual information (i.e. school report cards that they may never have received or may not be able to understand).
- In the **endline** survey, researchers asked parents again about *beliefs* about their child's academic performance by asking them to imagine a new test taken that day. These **endline beliefs** are also measured out of 100 points.

Next are STATA summaries of key variables, separately for the control group and the treatment group.

Control group:

Overall Score

Percentiles		Smallest		
1%	9	0		
5%	18	2		
10%	24	2	Obs	2,654
25%	35	2	Sum of Wgt.	2,654
50%	47		Mean	47.13075
		Largest	Std. Dev.	17.44873
75%	59	98		
90%	70	98	Variance	304.4582
95%	76	99	Skewness	.0712081
99%	88	100	Kurtosis	2.701315

Treatment group:

Overall Score

Percentiles		Smallest		
1%	10	0		
5%	18	0		
10%	24	0	Obs	2,614
25%	34	0	Sum of Wgt.	2,614
50%	46		Mean	46.35731
		Largest	Std. Dev.	17.53062
75%	58	95		
90%	70	96	Variance	307.3227
95%	76	97	Skewness	.1070127
99%	88	100	Kurtosis	2.651868

Supplement for Question (5), continues on the next page >>>>>

Supplement for Question (5), cont'd:

Control group:

Endline Beliefs (out of 100 points)

Percentiles		Smallest		
1%	20	0		
5%	35	1		
10%	40	2	Obs	2,642
25%	50	4	Sum of Wgt.	2,642
50%	65		Mean	63.56283
		Largest	Std. Dev.	17.65563
75%	75	100		
90%	85	100	Variance	311.7213
95%	90	100	Skewness	1.28091
99%	95	**350**	Kurtosis	28.7242

Note: There is an outlier in the posted replication data, which is shown in boldface above.

Treatment group:

Endline Beliefs (out of 100 points)

Percentiles		Smallest		
1%	15	0		
5%	25	0		
10%	30	0	Obs	2,602
25%	45	1	Sum of Wgt.	2,602
50%	55		Mean	56.14105
		Largest	Std. Dev.	18.44784
75%	70	100		
90%	80	100	Variance	340.3227
95%	85	100	Skewness	-.1299682
99%	95	100	Kurtosis	2.635569

Supplement for Question (5), continues on the next page >>>>>

Supplement for Question (5), cont'd:

Next, consider the excerpt below.

> **Excerpt (p. 21):** I now examine whether information changes beliefs by looking at the impact of information on mean beliefs measured at endline. Recall that, unlike beliefs measured at baseline, the beliefs question asked at endline was not asking about last-term test scores; instead, it asked how well parents thought their child would do on a hypothetical test taken that same day. The prediction is thus that providing information should decrease the gap between parents' endline beliefs and their child's last-term scores. Information cuts the gap nearly in half.

Table 1 below comes from Tables 2 and C.7 in the original paper as well as direct analysis of the replication data.

Table 1: Information Treatment Effects

Dependent variable: Endline Beliefs (out of 100 points)			
	(1)	(2)	(3)
Explanatory variables:			
Treat × Overall Score	0.405 (0.024)	0.408 (0.023)	0.408 (0.023)
Overall Score	0.305 (0.017)	0.303 (0.017)	0.303 (0.017)
Treat	-25.998 (1.208)	-26.023 (1.167)	-26.023 (1.167)
Dummy (HH=968, Ref=2)	-	-	284.147 (14.812)
Constant	49.188 (0.859)	49.213 (0.830)	49.213 (0.830)
Observations	5,244	5,243	5,244
R-squared	0.3095	0.3228	0.3548

Notes: Shows OLS results for three regressions. Data sources are baseline survey, baseline test score data, and endline survey data. Each observation is a child. The dependent variable is the parent's endline beliefs about the child's performance (out of 100 points) on a hypothetical test taken the same day as the endline survey. *Overall score* is the child's actual average overall score (out of 100 points) on the subjects of math, English, and Chichewa. *Treat* is a dummy that equals 1 for those children in the treatment group and equals 0 otherwise. Standard errors are in parentheses. Regression (1) includes an outlier: the observation for Household id=968 and reference child=2 with endline beliefs of 350, which is outside the possible range of values (0 to 100). Regression (2) excludes that outlier. Regression (3) keeps that outlier but includes a dummy variable for it.