ECO220Y1Y, Test #4, Prof. Murdock

March 8, 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 4 questions (some 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 table, and Student t table) 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.
- ***Unless otherwise specified, you choose the significance level. Absent any special considerations, you may choose α = 0.05.***
- ***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) [16 pts] A researcher wishes to prove that less than 40% of students support the changes announced by the Ford government in January 2019 to tuition, the Ontario Student Assistance Program, and student fees. If 32% of all students support the changes, what is the chance that a random sample of 250 students provides insufficient proof to meet a 5% significance level? In other words, what is the probability of a Type II error? <u>Answer with hypotheses in formal notation, TWO fully-labelled graphs, a quantitative analysis & the requested probability.</u>

(2) See ***Supplement for Question (2)***: *The 2018 World Happiness Report.*

(a) [9 pts] See **Figure 2.2**. China is ranked 86. For 2015-2017 combined, the sample size for China is 12,779. The exact values of the 95% confidence region shown in the figure are: [5.197142527, 5.29445751]. What is the value of the margin of error (ME)? Next, using that ME as a starting point, what is the *sample standard deviation* of happiness across people in the sample from China? Answer with a quantitative analysis.

(b) [5 pts] See **Table A7**. What are the hypotheses to test whether those with no family abroad *differ* from those with family abroad for ***life evaluations***? What is the *formula* for the test statistic? (Do NOT compute the test statistic.) Answer with hypotheses in formal notation & the relevant test statistic formula.

(c) [5 pts] See **Table A7**. What are the hypotheses to test whether those with no family abroad *differ* from those with family abroad for ***smiled yesterday***? (See the notes below the table.) What is the *formula* for the test statistic? (Do NOT compute the test statistic.) Answer with hypotheses in formal notation & the relevant test statistic formula.

(d) [5 pts] For **Regression #2**, what is the *interpretation* of 0.077? Answer with 1 – 2 sentences.

(e) [5 pts] For **Regression #3**, what is the *interpretation* of 0.853? Answer with 1 – 2 sentences.

(f) [3 pts] For **Regression #1**, which is the histogram of the residuals? Answer by writing A, B, C, D, E, or F in the blank.

The histogram of the residuals for Regression #1 must be Histogram **____**.

(3) See ***Supplement for Question (3)***: *Parents' Beliefs About Their Children's Academic Ability*.

(a) [8 pts] See **Table 1**. What is the coefficient of correlation between *believed math scores* and *believed English scores*? Answer with a quantitative analysis.

(b) [12 pts] See **Table 1**. Considering *academic performance*, is the *difference* between mean math and English scores *statistically significant*? (Do NOT answer Part **(c)** here.) Answer with the correct approach, analysis & 1 sentence.

(c) [5 pts] See **Table 1**. Considering *academic performance*, is the *point estimate* of the *difference* between mean math and English scores *economically significant*? Explain why or why not. Answer with 1 – 2 sentences.

(d) [4 pts] See the **STATA regression**. What is the missing value for "`Root MSE`"? <u>Answer with a quantitative analysis.</u>

(e) [5 pts] See the **STATA regression**. What are the two missing values under "`[95% Conf. Interval]`"? <u>Answer with a quantitative analysis.</u>

(4) See ***Supplement for Question (4)***: *ADHD Diagnoses.*

(a) [3 pts] For all birth months combined, what is the *overall rate per 10,000 children* of ADHD diagnoses in these data? <u>Answer with a quantitative analysis.</u>

(b) [15 pts] Comparing the youngest sixth of children (those born in July or August) with the oldest sixth of children (those born in September or October), how strong is the evidence that the rate of ADHD diagnoses is higher for the younger children? Answer with hypotheses in formal notation, a quantitative analysis, a P-value & 1 – 2 sentences.

This *Supplement* contains the aid sheets (formulas, Normal table, and Student t table) 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:** $r = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^{n} z_{x_i} z_{y_i}}{n-1}$

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

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

<u>Laws of expected value</u>:

$E[c] = c$

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

$E[cX] = cE[X]$

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

<u>Laws of variance</u>:

$V[c] = 0$

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

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

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

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

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

<u>Laws of covariance</u>:

$COV[X,c] = 0$

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

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

<u>Sampling distribution of $\bar{X}$</u>:

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

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

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

<u>Sampling distribution of $\hat{P}$</u>:

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

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

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

<u>Sampling distribution of $(\hat{P}_2 - \hat{P}_1)$</u>:

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

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

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

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

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

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

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

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

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

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

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

<u>Inference about a population proportion</u>:

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

<u>Inference about comparing two population proportions</u>:

***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:** $\bar{X} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$ **DoF (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:** $(\bar{X}_1-\bar{X}_2) \pm t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$ **DoF:** $\nu = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{s_1^2}{n_1}\right)^2+\frac{1}{n_2-1}\left(\frac{s_2^2}{n_2}\right)^2}$

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:** $(\bar{X}_1-\bar{X}_2) \pm t_{\alpha/2}\sqrt{\frac{s_p^2}{n_1}+\frac{s_p^2}{n_2}}$ **DoF:** $\nu = n_1+n_2-2$

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:** $\bar{X}_d \pm t_{\alpha/2}\frac{s_d}{\sqrt{n}}$ **DoF:** $\nu = n-1$

SIMPLE REGRESSION:

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

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

$s_y^2 = \frac{SST}{n-1}$ **Coefficient of determination:** $R^2 = (r)^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$ **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:** $b_1 \pm t_{\alpha/2}s.e.(b_1)$ **Standard error of slope:** $s.e.(b_1) = s_{b_1} = \frac{s_e}{\sqrt{(n-1)s_x^2}}$

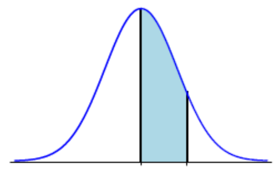
DoF: $\nu = n-2$

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

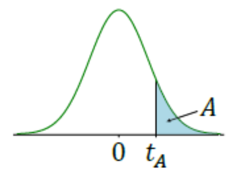
Confidence interval for predicted mean at given value of x (x_g):

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



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

Supplement for Question (2): 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)

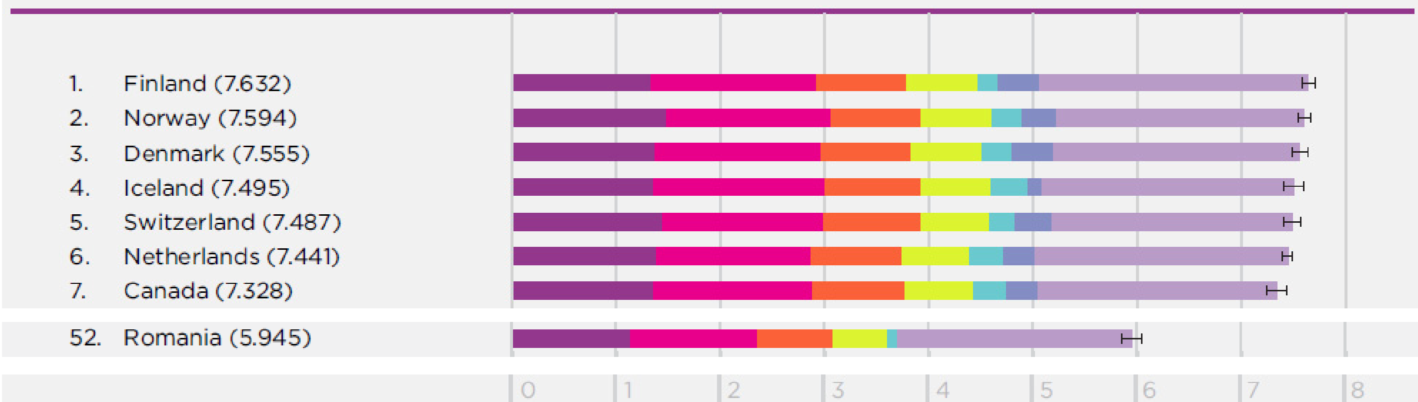


Figure 2.2: Ranking of Happiness 2015–2017 (Part 2)

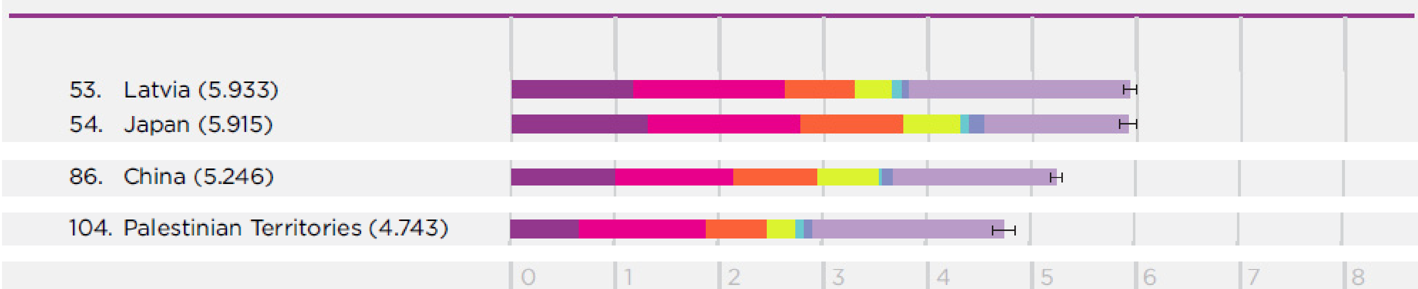
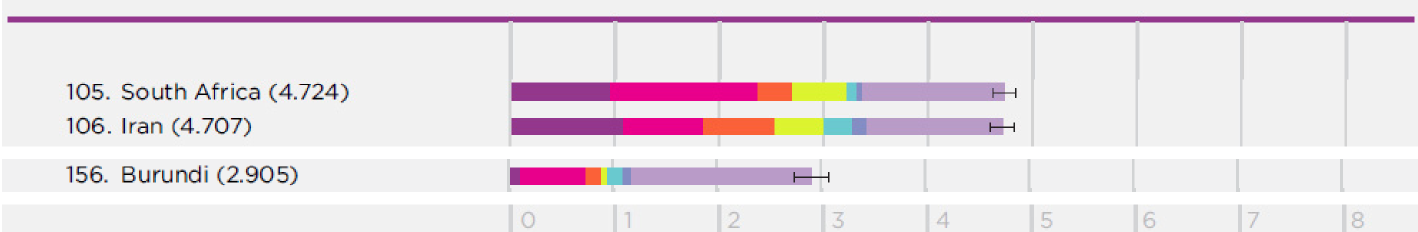


Figure 2.2: Ranking of Happiness 2015–2017 (Part 3)



Supplement for Question (2), continues on the next page >>>>>

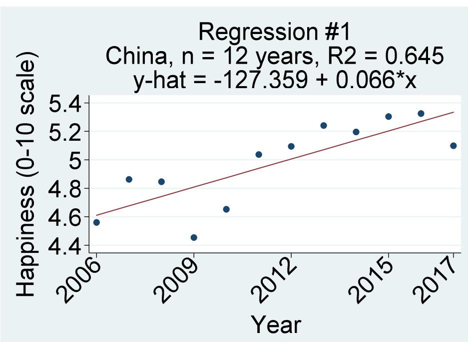
Supplement for Question (2), cont'd: Below is an excerpt of Table A7 on p. 109.

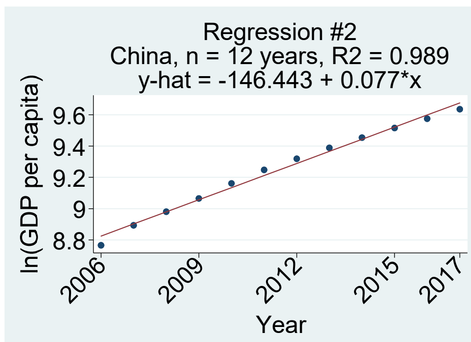
Table A7: Summary Statistics for Respondents with and Without Relative Abroad

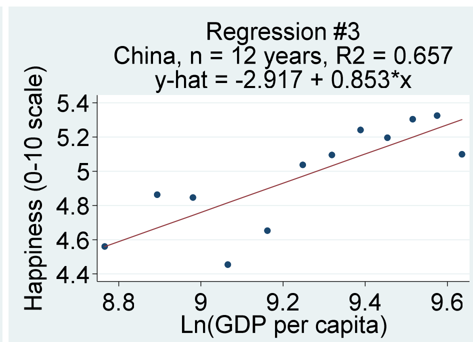
	No family abroad, N=19,933		Family abroad, N=3,976	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
Live evaluations (0-10 scale)	6.414	2.305	6.336	2.287
Smiled yesterday (1=yes)	0.859	0.348	0.868	0.338

Notes: The sample 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. Smiled yesterday is a binary indicator for whether the respondent reported smiling the previous day.

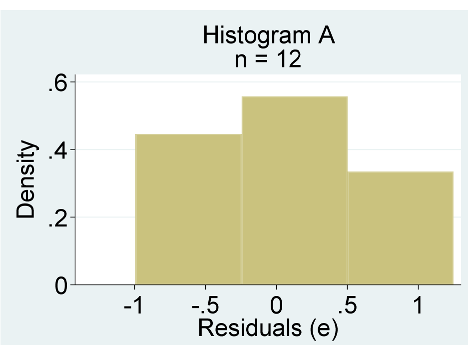
The data include annual observations for China for 2006 through 2017. On page 20, the report explains the GDP per capita measure as "Purchasing Power Parity (PPP) adjusted to constant 2011 international dollars, taken from the World Development Indicators (WDI) released by the World Bank in September 2017." See Regressions #1, #2 and #3 below.

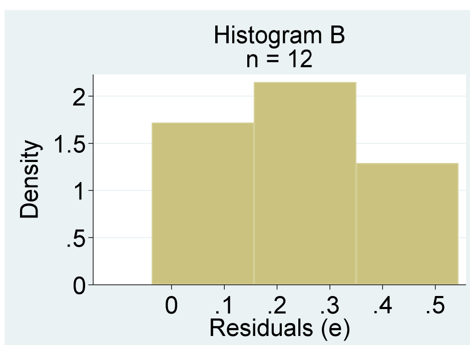


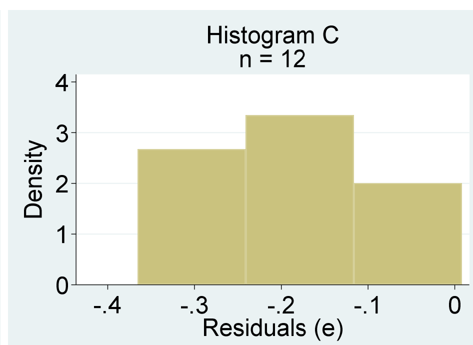


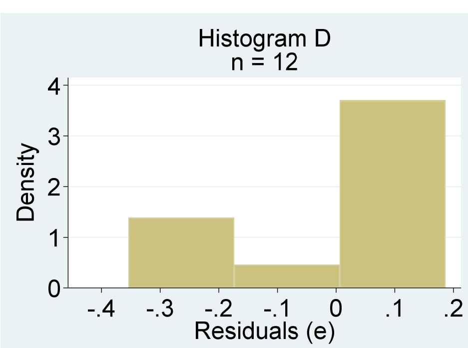


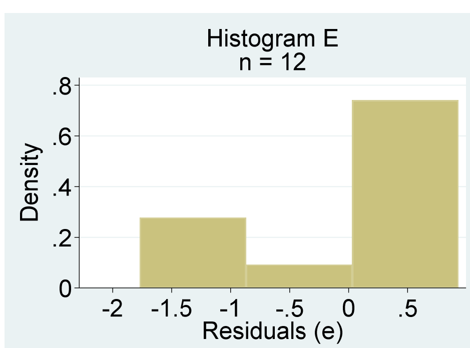
Below are the histogram choices for Part **(f)** of Question **(2)**.

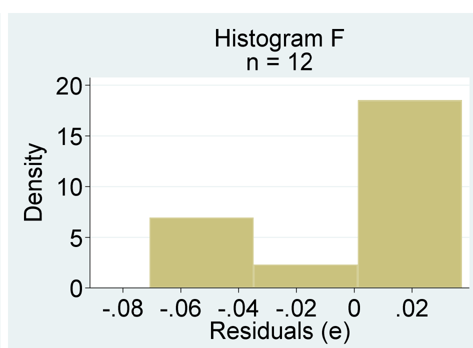












Supplement for Question (3): In the article "Parents' Beliefs About Their Children's Academic Ability: Implications for Educational Investments," Dizon-Ross (2019) shows how providing parents with clear, understandable information about their child's academic progress can help parents make more informed decisions. She does a field experiment with 5,268 children from 39 randomly selected primary schools in two districts (Machinga and Balaka) in Malawi. Consider the table below, which is an excerpt from that paper. It reports some basic summary statistics for the sample. All scores are out of 100 possible points.

Table 1. Baseline summary statistics

	Mean	SD
Academic Performance (Average Achievement Scores)		
Overall score	46.8	17.5
Math score	44.9	20.2
English score	44.2	20.1
Chichewa score	51.2	22.5
(Math – English) score	0.71	19.5
Respondent's Beliefs about Child's Academic Performance		
Believed Overall score	62.4	16.5
Believed Math score	64.7	19.0
Believed English score	55.3	20.9
Believed Chichewa score	66.8	19.4
Beliefs about (Math – English) score	9.48	21.5
Sample size (number of kids)	5,268	

Source: Excerpt from Table 1 on page 19 of Dizon-Ross (2019) https://www.aeaweb.org/articles?id=10.1257/aer.20171172.

STATA regression: The Stata regression output below shows a simple regression where the y-variable is academic performance measured by the overall score (overall). The x-variable is the respondent's beliefs about the child's overall academic score (b_overall). The sample size is slightly smaller than Table 1 because the data contain some missing values for beliefs. I have intentionally erased some numbers below. The specific values that you are asked to find are highlighted in grey (__________).

```
. regress overall b_overall;

      Source |       SS           df       MS      Number of obs   =     5,256
-------------+----------------------------------   F(1, 5254)      =
       Model |  187220.818         1               Prob > F        =
    Residual |  1418634.27     5,254               R-squared       =
-------------+----------------------------------   Adj R-squared   =
       Total |  1605855.09     5,255               Root MSE        =  _________

------------------------------------------------------------------------------
     overall |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   b_overall |   .3615425   .0137301    26.33   0.000     _________  _________
       _cons |   24.18992   .8856223    27.31   0.000
------------------------------------------------------------------------------
```

Supplement for Question (4): In the article "Attention Deficit-Hyperactivity Disorder and Month of School Enrollment," Layton et al. (2018) study the rate of claims-based ADHD diagnosis (insurance claims). Below are excerpts.

Excerpt from PowerPoint slides that accompany the article:

- In U.S. states that have an age cutoff at September 1 for entry to kindergarten, the rates of diagnosis and treatment of ADHD were higher among children born in August than among those born in September, which suggests that the age within a class cohort influences diagnosis.
- Data came from a large insurance database.

Excerpt from p. 2,122 of the article: Younger children in a school grade cohort may be more likely to receive a diagnosis of attention deficit-hyperactivity disorder (ADHD) than their older peers because of age-based variation in behavior that may be attributed to ADHD rather than to the younger age of the children. Most U.S. states have arbitrary age cutoffs for entry into public school. Therefore, within the same grade, children with birthdays close to the cutoff date can differ in age by nearly 1 year.

We used data from 2007 through 2015 from a large insurance database to compare the rate of ADHD diagnosis among children born in August with that among children born in September.

The study included 407,846 children in U.S. states [that maintained a September 1 cutoff in 2010 through 2014] who were born in the period from 2007 through 2009 and were followed through December 2015.

Figure 1. Differences in Diagnosis Rates of Attention Deficit-Hyperactivity Disorder (ADHD) According to Month of Birth

	Jan.	**Feb.**	**Mar.**	**Apr.**	**May**	**Jun.**	**Jul.**	**Aug.**	**Sep.**	**Oct.**	**Nov.**	**Dec.**	**Total**
Total number of children	32,690	31,238	34,405	34,565	34,977	34,415	36,577	36,319	35,353	34,405	31,285	31,617	407,846
Number of children with ADHD	265	280	307	312	287	317	320	309	225	240	232	243	3,337
Rate per 10,000 children	81.1	89.6	89.2	90.3	82.1	92.1	87.5	85.1	63.6	69.8	74.2	76.9	