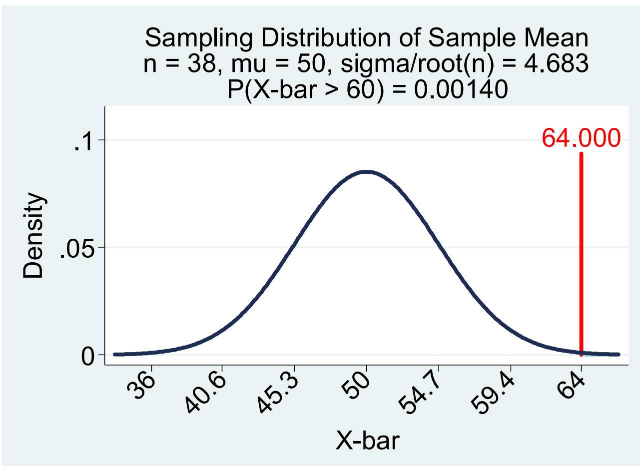
ECO220Y1Y, APRIL 2017, FINAL EXAM: SOLUTIONS

(1) Define the random variable X to be a student's percentile result. $X \sim U[0, 100]$ and $E[X] = \frac{a+b}{2} = \frac{0+100}{2} = 50$ and $V[X] = \frac{(b-a)^2}{12} = \frac{(100-0)^2}{12} = 833.33$ and $SD[X] = \sqrt{833.33} = 28.87$.

Given that the population is not too far from Normal (it is Uniform), a sample size of 38 is certainly sufficiently large to employ the Central Limit Theorem: the distribution of $\bar{X}$ will be Normal even though X is not.

Hence, $\bar{X} \sim N(50, 21.93)$ with $E[\bar{X}] = E[X] = 50$ and $V[\bar{X}] = \frac{V[X]}{n} = \frac{833.33}{38} = 21.93$ and $SD[\bar{X}] = \sqrt{21.93} = 4.683$.

$P(\bar{X} > 64) = P\left(Z > \frac{64-50}{4.683}\right) = P(Z > 2.99) = 0.5 - 0.4986 = 0.0014$ (Note: It does not matter if the inequality is written as a strong or weak inequality because X-bar is a continuous random variable.)



(2) (a) Start with the intercept coefficient of 9.818475368: According to the OLS line, in June 2005 the predicted price of a Big Mac hamburger is 9.82 Chinese Yuan (which is a bit below the actual price of 10.50 Chinese Yuan). Next, the slope coefficient of 0.06467537: On average the price of a Big Mac hamburger in China has gone up by 0.065 Chinese Yuan per month during the period from June 2005 and January 2017.

(b) $\hat{P}_t = 9.82 + 0.78t$ with the R^2 unchanged at 0.98. The results show that prices on average are increasing by 0.065 ¥ per month, so that means they are increasing by 0.78 ¥ per year: the coefficient on t would be 0.78 if t were measured in years. The constant term is unaffected because zero years is the same a zero months since June 2005. The R^2 is a unit-free statistic and it would *not* be affected by a change in the units of measurement of the x-variable from months to years (or any other change in the units of measurement of the x-variable and/or the y-variable).

(c) The number 4.30887E-17 is the P-value for the test of the overall statistical significance of the model and it is extremely tiny indicating an extremely statistically significant result. This is not surprising because the R-squared is near perfect: there is very little scatter and the price is rising steadily in a linear fashion. Hence, these data allow us to decisively reject the possibility that Big Mac prices have been constant (not rising) during this period in China.

(d) Yes, that would clearly create a significant outlier: that would be nearly double the price in the earliest time period and would substantially affect the OLS results as it bucks the trend of increasing prices over time (as prices are much lower than 19.50 in the next time periods). The number 0.448771056 is a measure of the standard deviation of the residuals (also called the Root MSE or s_e) and it would increase substantially with this outlier. The s_e is measured in the same units as the y-variable, which is Chinese Yuan, and it measures the amount of scatter around the OLS line: the outlier would greatly increase the amount of scatter. Relative to the double-digit prices of Big Macs – ranging from 10.50 ¥ to 19.60 ¥ – 0.45 ¥ is quite small amount of scatter, which is consistent with the very high R-squared values reported.

(3) (a) For every 1 percent increase in the cumulative number of cars produced at an automobile manufacturing plant in the mid-2000s, we observe that the average number of defects per car decreases by 0.3 percent on average.

(b) There are four regressions reported in Table 1. Column (1), Panel B corresponds with Figure 2.

(c) About 155,000 cars = exp(11.95), where 11.95 is an approximation obtained from Figure 2.

(d) Use the confidence interval estimator formula: $b_j \pm t_{\alpha/2} s_{b_j}$ with $\nu = n - k - 1$ to obtain $-0.336 \pm 2.692 * 0.017$ with $\nu = 47 - 2 - 1 = 44$, which yields a LCL of -0.382 and a UCL of -0.290.

(e) $H_0: \gamma = 0$ versus $H_1: \gamma \neq 0$. The test statistic is $t = \frac{0.007-0}{0.002} = 3.5$ with $\nu = n - 2 - 1 = 44$. The critical values for this two-tailed test are -2.015 and 2.015.

(f) Figure 1 in the *Supplement* clearly shows that average defects per car are *decreasing* over time. The positive coefficient on the time trend in Table 1 simply means that *after controlling for* cumulative production there is a slight positive increase in defects over time. If we were to drop cumulative production, the time trend coefficient would become a substantial *negative* number. The key point is that the important variable in determining the rate of defects is experience as measured by units produced and *not* simply calendar time passing.

(4) (a) A yes/no question about whether or not there is a difference requires a hypothesis testing approach to statistical inference: there is no direction specified so we must do a two-tailed test.

$H_0: p_C - p_T = 0$

$H_1: p_C - p_T \neq 0$

$\hat{P}_C = 0.548$

$$\hat{P}_T = \frac{0.521 * 2{,}120 + 0.536 * 1{,}867 + 0.525 * 2{,}063}{2{,}120 + 1{,}867 + 2{,}063} = \frac{3{,}188}{6{,}050} = 0.527$$

$\bar{P} = \frac{X_C + X_T}{n_C + n_T} = \frac{0.548*2{,}140+3{,}188}{2{,}140+6{,}050} = 0.53244$

$$z = \frac{\hat{P}_C - \hat{P}_T}{\sqrt{\frac{\bar{P}(1-\bar{P})}{n_C} + \frac{\bar{P}(1-\bar{P})}{n_T}}} = \frac{0.548 - 0.527}{\sqrt{\frac{0.53244(1-0.53244)}{2{,}140} + \frac{0.53244(1-0.53244)}{6{,}050}}} = \frac{0.021}{0.01255} = 1.67$$

$$P - value = P(Z < -1.67) + P(Z > 1.67) = 2 * (0.5 - 0.4525) = 0.095$$

These results are marginally statistically significant – just meet a 10% significance level, but do not meet a 5% significance level – and not economically significant. There is a small difference in the proportion of flights that are efficient between the control and treatment groups: a 2.1 percentage point difference between the control group (54.8% efficient) and the treatment group (52.7% efficient). Also, the direction of these results are not even what we may have expected: the control group that got no feedback on their performance actually did slightly better than the treatment groups that got individual feedback. Further, the P-value is fairly large, which means this difference could be caused by sampling error even if there were absolutely no difference in the population proportions: it is not statistically significant at the usual 5% significance level.

(b) According to Table 4, the percent of ALL flights (i.e. flights by any of the captains) that were efficient is <u>31.2</u> percent before the experiment versus <u>53.3</u> percent after the experiment. This is a <u>22.1</u> percentage point increase and a <u>70.8</u> percent increase.

(5) Let X be the random variable that records the API for a randomly selected day. The question asks us to find $P(100 < X < 120)$. [Note: It does not matter whether these are written as weak or strong inequalities because X is a continuous random variable.]

$$P(100 < X < 120) = P\left(\frac{100 - 75}{20} < Z < \frac{120 - 75}{20}\right) = P(1.25 < Z < 2.25) = 0.4878 - 0.3944 = 0.0934$$

(6) Let $X_1, X_2, \ldots, X_{10}$ be the random variables that record the money raised each day. The question asks us to find $SD[X_1 + X_2 + \cdots + X_{10}] = \sqrt{V[X_1] + V[X_2] + \cdots + V[X_{10}]} = \sqrt{10 * 15{,}000^2} = \$47{,}434.$

(7) (a) Males aged 5-19 years

(b) $\widehat{M}_{it} = 0.86 + 0.006\, PP_{it} + 0.45\, YR1990_{it} + 0.009\, YR1990_{it} * PP_{it}$

(c) For females 50 years old and older, mortality rates have dropped for all counties – ranging from rich to poor – between 1990 and 2010. However, mortality inequality has worsened because mortality rates have dropped less for females in the poorest counties than in the richest counties. In other words, because the reduction in mortality rates has been unevenly distributed, in 2010 there is a bigger discrepancy in mortality rates between the richest and poorest counties than there was in 1990. The increased steepness of the slope – remember steeper means more inequality – is statistically significant at the 5% level. It is also economically significant in that for each decile increase in the poverty percentile we see an extra 1.0 deaths per 1,000 in 1990 versus an extra 1.6 deaths per 1,000 in 2010. Hence, unlike younger females or the males where the amount of inequality is decreasing or about the same, for older females health inequality between the rich and poor is increasing.

(8) (a) No, the event that one defaults is definitely not disjoint (mutually exclusive) from an event where the second one defaults: they can *both* default. While they are independent events, as implied by "perfectly uncorrelated," they are not mutually exclusive events.

(b) This is a Binomial probability problem. Define X to be a random variable that counts the number of mortgages that default.

$$P(X = x) = \frac{n!}{x!\,(n-x)!} p^x (1-p)^{n-x}$$

Alpha Pool: $P(X = 5) = \frac{5!}{5!(5-5)!} 0.05^5 (1 - 0.05)^0 = 0.000000313 = 0.00003\%$

Beta Pool: $P(X = 4) = \frac{5!}{4!(5-4)!} 0.05^4 (1 - 0.05)^1 = 0.000029688 = 0.003\%$

Gamma Pool: $P(X = 3) = \frac{5!}{3!(5-3)!} 0.05^3 (1 - 0.05)^2 = 0.001128125 = 0.1\%$

Delta Pool: $P(X = 2) = \frac{5!}{2!(5-2)!} 0.05^2 (1 - 0.05)^3 = 0.021434375 = 2.1\%$

Epsilon Pool: $P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{5!}{0!(5-0)!} 0.05^0 (1 - 0.05)^5 = 0.226219063 = 22.6\%$ (NOTE: This is *not* the same as $P(X = 1)$.)