# *Spring 2018 - Review Problems for the Final - Solutions*

**1.** Chapter 26, exercise set F, problems 4 and 5.

*See solutions at the back of the book.*

**2.** Every day, the quality control engineer for ACME Dairies randomly selects 25 half-gallon (64 fl oz) cartons of whole milk from the day's production run and carefully measures the quantity of milk in each one. If the average amount of milk in this sample differs *significantly* from 64 fl oz, at the 1% significance level, she recalibrates the carton-filling apparatus.

(a) What does '*...significantly... at the* 1% *significance level*' mean?

*This is statistical significance. In this case, it means that assuming the null hypothesis, the probability is* 1% *or less, that the difference between the sample average and the (null-hypothetical) expected average is as or more extreme than the observed difference.*

(b) State the null and alternative hypotheses for this test in terms of the appropriate parameter.

$H_0$*: The average amount of milk in the daily run of half-gallon cartons is* 64 *ounces.*

$H_1$ *: The average amount of milk in the daily run of half-gallon cartons is* ***not*** 64 *ounces.*

*(A* ***two-tailed*** *test —* $H_1$ *does not specify whether the average is greater than or less than* 64*, just that it is different than* 64*.)*

(c) Today's sample of 25 cartons has an average of 64.21 fl oz with a standard deviation of 0.37 fl oz. What is the test statistic? What probability distribution does it follow? What is the p-value?

*The test statistic here is*

$$t^* = \frac{(\textit{observed average}) - (H_0\textit{-expected average})}{SE(\textit{average})} = \frac{64.21 - 64}{SD^+/\sqrt{25}} \approx \frac{0.21}{0.378/5} \approx 2.78$$

**Comments:** *The sample size* $n = 25$ *is relatively small and we do not know the SD of the error-box, so we approximate it with the* $SD^+$ *of the sample. The test statistic follows the t-distribution with* $25 - 1 = 24$ *degrees of freedom.*

*The p-value is estimated by looking in the row for* 24 *degrees of freedom in the t-table in the textbook. The area to the right of* $t^* = 2.78$ *is between* 1% *and* 0.5% *(because* $2.49 < 2.78 < 2.80$*). This is a* ***two-tailed*** *test, because we don't have an expected direction for the difference — a priori, it could be positive or negative. This means that the p-value is two times the entry in the table, i.e.,*

$$1\% < p^* < 2\%.$$

*(Using more precise tools, you would find that* $p^* \approx 1.04\%$*.)*

(d) What do you conclude?

*Since* $p^* > 1\%$ *(though it's very close!), the ACME Dairies protocol says that the machines do* ***not*** *need to be recalibrated.*

(e) What additional assumptions, if any, are needed to justify the methods (and conclusions) of this test of significance?

*To justify the use of the t-test, we must assume that the error-box associated with the filling of the milk cartons follows the normal curve (once the errors are converted to standard units).*

**3.** A marketing firm wants to survey a sample of adults from a large state to estimate the percentage of adults in the state who prefer streaming movies to watching them on cable TV. Describe the advantages and disadvantages of each of the following sample types that they might use: (a) a *simple random sample*, (b) a *convenience sample* or (c) a *multistage cluster sample.*

*(a) A simple random sample, has the smallest standard errors, but these are difficult to produce in practice from large populations. (b) A convenience sample is easy to collect (hence the name), but there is usually no way to produce meaningful standard errors or reliable confidence intervals. (c) A multistage cluster sample is the happy medium. They are more practical — they are easier to collect that simple random samples. and we can find meaningful standard errors and there appropriate box models that can be used to construct reliable confidence intervals, though the standard errors will be larger than in the case of a simple random sample (and so give less precise estimates).*

**4.** A researcher claims to have found a strong correlation ($r = 0.88$) between a person's blood alcohol content (BAC), one hour after drinking, and the type of alcohol they consume (beer, wine or hard liquor).

What is wrong with the researcher's claim? What would make more sense here? Explain.

*The correlation does not make sense because you can't compute the correlation between BAC (numerical) and alcohol type (nonnumerical). There is no way to interpret the number* 0.88 *(and it is not clear how it was computed).*

*The simplest correlation that would make sense here is a correlation between the amount of alcohol consumed and the BAC an hour later. If one wants to study the effect of the* ***type*** *of alcoholic beverage and BAC an hour later, a more involved study is needed.*

*E.g., the researcher could study the effect of each type of alcoholic beverage separately and* ***control for various confounding factors****, like the amount of alcohol that is consumed, weight and gender of the drinker, etc.*

**5.** In a calculus class with 180 students, the final exam score contributed 50% of the course score, the midterm score contributed 30% of the course score and the average homework score contributed 20% of the course score.

After the course was over, the instructor computed three correlation coefficients based on the class data:

- $r_1$ = correlation between average homework score and midterm exam score,
- $r_2$ = correlation between average homework score and final exam score,
- $r_3$ = correlation between average homework score and score in the class.

The three numbers she computed were 0.3521, 0.5582 and 0.4112, but she forgot to label them. Match each number with the appropriate correlation coefficient and explain your choices.

*Since the homework contributes directly to the score in the class, we expect $r_3$ to be the highest, i.e., $r_3 = 0.5582$. Since only about half the homework was completed by the time of the midterm, the correlation between the average of* all *the homework scores and the midterm is likely to be lower than the correlation between the average on the homework and the final exam, so $r_1 = 0.3521 < 0.4112 = r_2$.*

**6.** Investigators studied the relationship between screen time (measured in hours/day) and obesity (measured with *body-mass-index* BMI) in adults age 20 - 40. They surveyed 5178 U.S. adults in this age group, and generated the following summary statistics:

$$\begin{array}{lll} \overline{X} = 10 & SD_X = 6 & \\ \overline{Y} = 27 & SD_Y = 9 & r = 0.6 \end{array}$$

where $X$ = hours of screen-time per day, and $Y$ = BMI.

(a) Use the *regression method* (or regression equation) to estimate the BMI for U.S. adults, aged 20 - 40 who have 7 hours of screen time per day. *Show your work.*

*7 hours of screen time per day is* $(7-10)/6 = -0.5\ SD_x$ ***below*** *average, so the regression method predicts that the BMI of adults who have 7 hours of screen time per day will be*

$$0.5 \times r \times SD_y = 0.5 \times 0.6 \times 9 \approx 2.7 \textit{ units}$$

***below*** *the average BMI of* 27*. I.e., the predicted average BMI for these people is about* 24.3*.*

*Using the regression equation method, the slope coefficient is*

$$\beta_1 = r \times \frac{SD_y}{SD_x} = 0.6 \times \frac{9}{6} = 0.9,$$

*and the intercept coefficient is*

$$\beta_0 = \overline{Y} - \beta_1 \overline{X} = 27 - 0.9 \times 10 = 18,$$

*so the regression equation is*

$$Y = 18 + 0.9X,$$

*and the predicted average BMI for adults who have* 7 *hours of screen time per day is*

$$Y(7) = 18 + 0.9 \cdot 7 = 24.3.$$

(b) What is the predicted BMI of a 28-year old woman who has 15 hours of screen-time per day? Include a '*give-or-take*' number with your estimate. *Show your work.*

*The predicted BMI for this woman is the average BMI of all adults (age* 20 - 40*) who have* 15 *hours of screen time per day, which is*

$$Y(15) = 18 + 0.9 \cdot 15 = 31.5.$$

*The 'give-or-take' number is given by the SER (root mean square error of regression) for predicting BMI from hours of screen time per day, which is*

$$SER = \sqrt{1 - 0.6^2} \times SD_y = 7.2,$$

*so the BMI of an* ***individual*** *28 year old woman who has 15 hours of screen time per day is predicted to be* $31.5 \pm 7.2$*.*

(c) Ivan is a 24-year old Russian graduate student at UCLA, who has about 12 hours of screen-time per day. Is it reasonable to predict that his BMI is somewhere between 21.6 and 36, based on the given information? *Explain your answer.*

*This is not a question of whether the calculations were done correctly. The data was collected from U.S. adults aged 20 - 40, and cannot be used to predict the BMI for a non-U.S. man, since cultural differences may have an effect on BMI.*

**7.** John Smith is running for office. One week before the election, his campaign manager hires a Polling firm to survey likely voters. The firm surveyed a simple random sample of 2700 likely voters and found that 51% favor Smith. They also found that of the 1250 women in the survey, 54% favor Smith.

You may assume that the survey was based on a simple random sample, that the population is in the millions and that to win the office, the candidate needs to win more than 50% of the votes cast.

(a) What percentage of the men in the survey favor Smith?

$675 = 54\% \times 1250$ *of the women surveyed favored Smith, and a total of* $51\% \times 2700 = 1377$ *of the people surveyed favored him. So,* $1377 - 675 = 702$ *men surveyed favored Smith. A total of* $2700 - 1250 = 1450$ *men were surveyed, so*

$$\frac{702}{1450} \times 100\% \approx 48.41\%$$

*of the men surveyed favor Smith.*

(b) Compute 95% confidence intervals for the percentage of women who favor Smith, the percentage of men who favor Smith and the percentage of likely voters who favor Smith.

**Women:** *The observed percentage is 54%, and the standard error is*

$$SE_W = \frac{\sqrt{0.54 \times 0.46}}{\sqrt{1250}} \times 100\% \approx 1.41\%.$$

*The 95% confidence interval for the percentage of women who favor Smith is*

$$(54\% \pm 2SE_W) = (54\% \pm 2.82\%).$$

**Men:** *The observed percentage is 48.41%, and the standard error is*

$$SE_M = \frac{\sqrt{0.4841 \times 0.5159}}{\sqrt{1450}} \times 100\% \approx 1.31\%.$$

*The 95% confidence interval for the percentage of men who favor Smith is*

$$(48.41\% \pm 2SE_M) = (48.41\% \pm 2.62\%).$$

**All:** *The observed percentage is 51%, and the standard error is*

$$SE_A = \frac{\sqrt{0.51 \times 0.49}}{\sqrt{2700}} \times 100\% \approx 0.96\%.$$

*The 95% confidence interval for the percentage of all likely voters who favor Smith is*

$$(51\% \pm 2SE_A) = (51\% \pm 1.92\%).$$

(c) *It seems that Smith's campaign should focus its resources on swaying more* ***men*** *to vote for him because it is almost certain that he will win over* 50% *of the women's vote.*

**8.** As part of a class project, a statistics student at a large university (15,000 students — 9000 men and 6000 women), went to the central plaza of the campus at noon one day, approached 100 students and asked them where they went to high school. His sample included 51 women and 49 men. Is it likely that the student's sampling procedure was like taking a simple random sample? Justify your answer as precisely as possible (using numbers, probability, etc.).

*If the student's sampling procedure was like taking a simple random sample, then it was like drawing 100 tickets at random,* without *replacement from a 0-1 box of 15000 tickets, where 60% of the tickets are* $\boxed{1}$*s and 40% are* $\boxed{0}$*s. The question now becomes:*

> ***How likely is it to draw 49%*** $\boxed{1}$***s (or less) from a box with 60%*** $\boxed{1}$***s, in 100 random draws?***

*To answer this question, we use the* ***Normal Approximation****. The SD of the box is*

$$SD = \sqrt{0.6 \times 0.4} \approx 0.49,$$

*and the* $SE_{\%}$ *for 100 draws from this box is*

$$SE_{\%} = \frac{0.49}{\sqrt{100}} \cdot 100\% = 4.9\%.$$

*(Technically, the $SE_{\%}$ is **slightly** smaller, because the draws are done without replacement, but since there are 15000 tickets in the box and only 100 are drawn, the correction factor is very close to 1.)*

*According to the normal approximation, the probability of drawing 49% (or fewer) $\boxed{1}$s from this box with a simple random sample is about equal to the area under the normal curve to the left of*

$$z = \frac{49\% - 60\%}{4.9\%} \approx -2.24,$$

*which is approximately 1.25%.*

*To summarize, the probability that a simple random sample of students from this University would have 49% men (or fewer) is about 1.25%, and we can conclude that the student's sample in this case was almost certainly **not** a simple random sample.*

*Indeed, from the description, it is clear that this was a **sample of convenience** (and perhaps biased towards women).*

**9.** According to the 1999 census, the median household income in the city of San Diego was $46,500. In 2004, a high-end grocery chain hires a statistical research firm to corroborate their marketing consultant's claim that median household income has gone up since 1999. The research firm takes a simple random sample of 600 San Diego households and finds that 55% of the sample households have incomes above $46,500.

Was the consultant right? Frame your answer in terms of an appropriate test of significance.

*To answer the question, we use a test of significance.*

- ***Box model:*** *A 0-1 box with a $\boxed{1}$ for every household in San Diego (in 2004) with income above $46,500 and a $\boxed{0}$ for every household in San Diego (in 2004) with income below $46,500.*
- ***Null hypothesis:*** *The median income has not gone up since 1999...*
  $H_0 : \%_{SanD} = 50\%$

  ***Alternative hypothesis:*** *The median income has gone up since 1999...*
  $H_A : \%_{SanD} > 50\%$
  *where $\%_{SanD}$ is the percentage of households in San Diego with income over $46,500.*
- ***Data:*** *The sample percentage of households with incomes above $46,500 is 55%.*
- ***Test statistic:*** *The observed percentage is 55% and the null-hypothetical expected percentage is 50%. Furthermore, the SD of the null-hypothetical box is $\sqrt{1/2 \times 1/2} = 1/2$, so the standard error is $SE_{\%} = \frac{0.5}{\sqrt{600}} \times 100\% \approx 2.04\%$.*
  *Hence the test statistic is*

  $$z = \frac{\text{observed }\% - \text{expected }\%}{SE_{\%}} = \frac{55\% - 50\%}{2.04\%} \approx 2.45.$$
- ***P-value*** *(observed significance level): The P-value here is the area under the normal curve to the right of $z = 2.45$ which is about $\frac{100\% - 98.57\%}{2} \approx 0.715\%$.*
- ***Conclusion:*** *The P-value is very low (less than 1%), so we reject the null hypothesis and conclude that the consultant was right — the median income in 2004 was higher than $46,500.*

**10.** Suppose that a fair die is rolled 3 times.

**a.** What is the probability that a ⚀ is observed ***at least once***?

*The probability of* no ⚀*s in* 3 *rolls is*

$$\frac{5}{6}\cdot\frac{5}{6}\cdot\frac{5}{6}\approx 57.87\%$$

*so the probability of* ***at least one*** ⚀ *in three rolls is* $100\% - 57.87\% = 42.13\%$*.*

**b.** What is the probability that a ⚀ is observed ***exactly once***?

*If a* ⚀ *is observed exactly once, then...*
*... it occurs on roll one, but not on rolls two and three; or it occurs on roll two, but not on rolls one or three; or it occurs on roll three, but not on rolls one or two.*
*Each of these three possibilities has the same probability, namely*

$$\frac{1}{6}\cdot\frac{5}{6}\cdot\frac{5}{6}\approx 0.11574,$$

*and all three are* ***mutually exclusive*** *so the probability that exactly one of them occurs is*

$$\frac{1}{6}\cdot\frac{5}{6}\cdot\frac{5}{6}+\frac{1}{6}\cdot\frac{5}{6}\cdot\frac{5}{6}+\frac{1}{6}\cdot\frac{5}{6}\cdot\frac{5}{6}\approx 0.3472 = 34.72\%.$$

**c.** What is the probability of that the ***sum*** of the three rolls is 4 or 5?

*First, we need to find the different configurations of three dice that result in sums of* 4 *or* 5*.*
*(i) The only way to obtain a sum of* 4 *in three rolls is with one* ⚁ *and two* ⚀*s. This can occur in three ways:* ⚀⚀⚁, ⚀⚁⚀ *or* ⚁⚀⚀*. Each of these configurations has the same probability, namely* $\frac{1}{6}\cdot\frac{1}{6}\cdot\frac{1}{6} = 1/216$*, and they are all mutually exclusive, so the probability of a sum of* 4 *in three rolls is*

$$3\times\frac{1}{216}=\frac{1}{72}\approx 1.389\%.$$

*(ii) The only ways to obtain a sum of* 5 *in three rolls is with (a) two* ⚀*s and one* ⚂ *or (b) one* ⚀ *and two* ⚁*s. In other words, the only way to obtain a sum of* 5 *is with one of the configurations*

⚀⚀⚂, ⚀⚂⚀, ⚂⚀⚀, ⚀⚁⚁,⚁⚀⚁ *or* ⚁⚁⚀.

*Each of these six configurations has the same probability, namely* $\frac{1}{6}\cdot\frac{1}{6}\cdot\frac{1}{6} = 1/216$*, and they are all mutually exclusive so the probability of a sum of* 5 *in three rolls is*

$$6\times\frac{1}{216}=\frac{1}{36}\approx 2.778\%.$$

*Finally, since a sum of* 4 *and a sum of* 5 *are mutually exclusive, the probability of a sum of* 4 *or a sum of* 5 *in three rolls is*

$$\frac{1}{72}+\frac{1}{36}=\frac{3}{72}=\frac{1}{24}\approx 4.167\%.$$

**11.** Suppose that a fair die is rolled 600 times.

**a.** What is the expected number of ⚀s?

*Given that the die is fair, the probability of observing a* ⚀ *on any given roll is* 1/6*, and the* ***expected number*** *of* ⚀*s is therefore equal to*

$$\frac{1}{6}\cdot 600 = 100.$$

**b.** What is the probability that a ⚀ is observed between 95 and 105 times?

*The SD of the ‘*$\boxed{\cdot}$*-box’ is* $\sqrt{1/6 \times 5/6} \approx 0.373$*, and the SE for the* ***number*** *of* $\boxed{\cdot}$*s in 600 draws is* $SD \times \sqrt{600} \approx 9.129$*. By the normal approximation, the probability of observing between 95 and 105* $\boxed{\cdot}$*s in 600 draws is therefore approximately equal to the area under the normal curve between*

$$\frac{94.5 - 100}{9.129} \approx -0.60 \quad \textit{and} \quad \frac{105.5 - 100}{9.129} \approx 0.60$$

*which is 45.15%.*

**c.** What is the probability that more than 110 $\boxed{\cdot}$s are observed?

*Once again, we invoke the normal approximation and conclude that this probability is approximately equal to the area under the normal curve to the right of*

$$\frac{110.5 - 100}{9.129} \approx 1.15$$

*which is*

$$\frac{100\% - 74.99\%}{2} = 12.505\%.$$

**d.** What is the probability that the ***sum*** of the 600 rolls is between 2070 and 2130?

*Once again, we invoke the normal approximation, but this time it is for the sum of draws from the box* $\left\lfloor \boxed{1}\,\boxed{2}\,\boxed{3}\,\boxed{4}\,\boxed{5}\,\boxed{6} \right\rfloor$.

*The average of this box is* $(1+2+3+4+5+6)/6 = 3.5$ *and the SD of this box is*

$$\sqrt{\frac{(1-3.5)^2 + (2-3.5)^2 + (3-3.5)^2 + (4-3.5)^2 + (5-3.5)^2 + (6-3.5)^2}{6}} \approx 1.708.$$

*The expected value of the sum of* 600 *draws from this box is* $600 \times 3.5 = 2100$ *and the standard error for the sum of* 600 *draws is* $SE = \sqrt{600} \times SD \approx 41.833$*. Using the normal approximation, we find that*

$$P(2070 \leq \textit{sum of } 600 \textit{ rolls} \leq 2130) \approx \textit{area under NC between } \frac{2070 - 2100}{41.833} \textit{ and } \frac{2130 - 2100}{41.833}.$$

*Now,* $\frac{2070-2100}{41.833} \approx -0.717$ *and* $\frac{2130-2100}{41.833} \approx 0.717$*, so the probability we seek is about* 53% *(between the table entries for* 0.70 *and* 0.75*, closer to the one for* 0.70 *and rounded).*

**12.** There are about 25,000 high schools in the United States and each high school has a principal. These 25,000 high schools also employ a total of about one million teachers. As part of a national survey of education, a simple random sample of 625 high schools is chosen.

(a) In 505 of the sample high schools the principal has an advanced degree. If possible, find an approximate 95% confidence interval for the percentage of all 25,000 high school principals who have advanced degrees. If this is not possible, explain why not.

*The sample percentage of principals with advanced degrees is* $505/625 \times 100\% = 80.8\%$*. The sample SD in this case is* $\sqrt{0.808 \times 0.192}$*, so the Standard error (for percentage) is*

$$SE_{\%} \approx \frac{\sqrt{0.808 \times 0.192}}{\sqrt{625}} \times 100\% \approx 1.58\%.$$

*Hence a 95% confidence interval for the percentage of high schools whose principal has an advanced degree is*

$$80.8\% \pm 2 \times 1.58\% = 80.8\% \pm 3.16\% \quad \textit{or } (76.64\%, 83.96\%).$$

(b) As it turned out, there were 250,000 students enrolled in the 625 sample high schools described above. These 250,000 students spent an average of 10.7 hours per week on homework, with a standard deviation of 3.5 hours. If you can, find an approximate 95% confidence interval for the average number of hours per week spent on homework of all U.S. high school students. If you cannot, explain why not.

*The sample of 250,000 students in this hypothetical example is* ***not*** *a simple random sample of U.S. high school students — taking all of the students from a random sample of high schools is not the same thing as a random sample of students from the whole country — it is a cluster sample of students. The methods we have been using do not apply in this case, and we cannot find a 95% confidence interval using these methods.*

**13.** A researcher studying the media consumption habits of U.S. adults suspects that Californians watch more 'reality' shows than New Yorkers. To test this hypothesis, she surveys a simple random sample of 1225 Californians and a simple random sample of 1444 New Yorkers. The New Yorkers surveyed watched an average of 4.36 hours per week of 'reality' shows, with an SD of 1.8 hours per week. The Californians watched an average of 4.43 hours per week of 'reality' shows, with an SD of 1.7 hours per week.

(a) Formulate appropriate null and alternative hypotheses in terms of a box model to test the researcher's hypothesis at the 5% significance level.

*There are two boxes — the California box has one ticket for every adult in CA with the number of hours of reality TV that person watches each week and the New York box has one ticket for every adult in NY with the number of hours of reality TV that person watches each week.*

*The null hypothesis says that the Californians and New Yorkers watch the same amount of reality television, on average, and the alternative hypothesis say that the CA average is higher than the NY average.*

*I.e., if $\mu_C$ is the average number of hours/week that Californians watch reality shows and $\mu_N$ is the average number of hours/week that New Yorkers watch reality shows, then the researcher's hypotheses are..*

$H_0 : \mu_C = \mu_N$, *and*

$H_A : \mu_C > \mu_N$.

(b) Find the test statistic and the $P$-value.

*The test statistic is*

$$z^* = \frac{\bar{C} - \bar{N}}{\sqrt{SE_C^2 + SE_N^2}},$$

*where $\bar{C}$ and $\bar{N}$ are the observed averages for Californians and New Yorkers respectively, and $SE_C$ and $SE_N$ are the standard errors for Californians and New Yorkers respectively. Plugging in the given sample statistics, we find that*

$$\bar{C} = 4.43,\ SE_C = \frac{1.7}{\sqrt{1225}} \approx 0.0486,\ \bar{N} = 4.36 \textit{ and } SE_N = \frac{1.8}{\sqrt{1444}} \approx 0.0474,$$

*so*

$$z^* = \frac{4.43 - 4.36}{\sqrt{(0.0486)^2 + (0.0474)^2}} \approx 1.03.$$

*The p-value is equal to the area under the normal curve to the right of $z^* = 1.03$, which is about* 15%.

(c) Is the researcher right? In what sense? Explain.

*The data does not support the researcher's claim. With a P-value of* 15%*, the difference between the sample averages for Californians and New Yorkers can be explained reasonably by chance error.*

**14.** Chapter 27, Review problem 7.

**Comment:** *This problem is similar to the radiation-surgery example in section* 4 *of chapter* 27*.*

(a) *The observed difference between the rates of recidivism is (control - treatment)* $49.4\% - 48.3\% = 1.1\%$. *The* $SE_{\%}$ *for the control group is*

$$SE_c = \frac{\sqrt{0.494 \times 0.506}}{\sqrt{154}} \times 100\% \approx 4\%$$

*and the* $SE_{\%}$ *for the treatment group is*

$$SE_t = \frac{\sqrt{0.483 \times 0.517}}{\sqrt{592}} \times 100\% \approx 2\%.$$

*so the SE for the difference is*

$$SE_{\mathit{diff}} = \sqrt{(0.04)^2 + (0.02)^2} \times 100\% \approx 4.5\%.$$

*This means that the test statistic is*

$$z^* = \frac{1.1\%}{4.5\%} \approx 0.24,$$

*and the p-value (from the table) is about* 40%*.*

**Conclusion:** *The observed difference in recidivism rates can be explained by chance — the income support did not seem to have a benefit.*

(b) *The observed difference between the average number of weeks worked (control - treatment) is* $24.3 - 16.8 = 7.5$*. The SE for the control group is*

$$SE_c = \frac{17.3}{\sqrt{154}} \approx 1.394$$

*and the SE for the treatment group is*

$$SE_t = \frac{15.9}{\sqrt{592}} \approx 0.653$$

*so the SE for the difference is*

$$SE_{\mathit{diff}} = \sqrt{(1.394)^2 + (0.653)^2} \approx 1.54.$$

*The test statistic is*

$$z^* = \frac{7.5}{1.54} \approx 4.87$$

*so the p-value is effectively* 0%*.*

**Conclusion:** *The observed difference in weeks-worked between the control and treatment group is **not** due to chance error. The released prisoners who received income support tended to work less than those who did not. Perhaps this explains the failure of the program to reduce recidivism.*