Mat 1372 HW14

6.10 Legalization of marijuana, Part I. The General Social Survey asked 1,578 US residents: "Do you think the use of marijuana should be made legal, or not?" 61% of the respondents said it should be made legal.[13]

(a) Is 61% a sample statistic or a population parameter? Explain.

(b) Construct a 95% confidence interval for the proportion of US residents who think marijuana should be made legal, and interpret it in the context of the data.

(c) A critic points out that this 95% confidence interval is only accurate if the statistic follows a normal distribution, or if the normal model is a good approximation. Is this true for these data? Explain.

(d) A news piece on this survey's findings states, "Majority of Americans think marijuana should be legalized." Based on your confidence interval, is this news piece's statement justified?

Sol: (a) sample statistic since it is from a sample of 1578 US residents

(b) $\hat{p} = 61\%$ and $SE = \sqrt{\frac{p(1-p)}{n}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{1578}} = \sqrt{\frac{0.61 \cdot 0.39}{1578}} = 0.012$

$n = 1578$

The 95% confidence interval is

$$\hat{p} \pm 1.96 \times SE \to 0.61 \pm 1.96 \cdot 0.012 \to (0.586, 0.634)$$

We are 95% confident that approximately 58.6% to 63.4% of Americans support the idea.

(c) Independence: a simple random sample

Sucess-failure condition. $np \approx n\hat{p} = 1578 \times 0.61 = 962.58 \geq 10$

$n(1-p) \approx n(1-\hat{p}) = 1578 \times 0.39 = 615.42 \geq 10$

Both conditions are satified.

(d) Yes, the interval $(0.586, 0.634)$ is above 50%

6.11 National Health Plan, Part I. A *Kaiser Family Foundation* poll for US adults in 2019 found that 79% of Democrats, 55% of Independents, and 24% of Republicans supported a generic "National Health Plan". There were 347 Democrats, 298 Republicans, and 617 Independents surveyed.[14]

(a) A political pundit on TV claims that a majority of Independents support a National Health Plan. Do these data provide strong evidence to support this type of statement?

(b) Would you expect a confidence interval for the proportion of Independents who oppose the public option plan to include 0.5? Explain.

Sol (a) From Independent group, the sample size is $n = 617$. and the support rate is 55%

Prepare: $H_0: p = 0.5$ (no majority)

H_A $p \neq 0.5$ (might be majority or minority)

$\alpha = 0.05$, null value $p_0 = 0.5$

Check: Independence: random samples

Success-failure condition: $np_0 = 617 \cdot 0.5 = 308.5 > 10$

$n(1-p_0) = 617 \cdot 0.5 = 308.5 > 10$

Calculate: $SE_{p_0} = \sqrt{\frac{p(1-p)}{n}} \approx \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.5(0.5)}{617}} = 0.02$

$$Z = \frac{0.55 - p_0}{SE_{p_0}} = \frac{0.55-0.5}{0.02} = 2.5$$

$$\text{P-value} = 2 \times P(Z > 2.5) = 2 \times (1 - P(Z < 2.5))$$
$$= 2 \times (1 - 0.9938) = 2 \cdot (0.0062) = 0.0124$$

Conclude: p-value $< 0.05 \Rightarrow$ We reject H_0 and since the support rate is 55%, so we will say TV's claim seems reasonable

(b) NO. Generally we expect a hypothesis test and confidence interval to align. so 0.5 won't be in the CI as well.

6.12 Is college worth it? Part I. Among a simple random sample of 331 American adults who do not have a four-year college degree and are not currently enrolled in school, 48% said they decided not to go to college because they could not afford school.[15]

(a) A newspaper article states that only a minority of the Americans who decide not to go to college do so because they cannot afford it and uses the point estimate from this survey as evidence. Conduct a hypothesis test to determine if these data provide strong evidence supporting this statement.

(b) Would you expect a confidence interval for the proportion of American adults who decide not to go to college because they cannot afford it to include 0.5? Explain.

Sol: (a) Sample Size $n = 331$, $\hat{p} = 48\%$ (not going to college due to the cost)

Prepare: $H_0: p = 0.5$ (no majority)

$H_A: p \neq 0.5$ (might be a majority or minority)

$\alpha = 0.05$, null value $p_0 = 0.5$

Check: Independence: a simple random sample

success-failure condition: $np = np_0 = 331 \cdot 0.5 > 10$

$n(1-p) = n(1-p_0) = 331 \cdot 0.5 > 10$

this sample proportion follows normal distribution

Calculate: $SE_{p_0} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.5 \cdot 0.5}{331}} = 0.027$

$$Z = \frac{0.48 - p_0}{SE_{p_0}} = \frac{0.48 - 0.5}{0.027} = -0.73$$

$\text{P-value} = 2 \cdot P(Z < -0.73) = 2 \cdot 0.2327 = 0.5654$

Conclude: P-value $> \alpha = 0.05$, We fail to reject H_0. The data didn't provide the strong evidence of that claim.

(b) Yes, since we fail to reject H_0

6.14 Is college worth it? Part II. Exercise 6.12 presents the results of a poll where 48% of 331 Americans who decide to not go to college do so because they cannot afford it.

(a) Calculate a 90% confidence interval for the proportion of Americans who decide to not go to college because they cannot afford it, and interpret the interval in context.

(b) Suppose we wanted the margin of error for the 90% confidence level to be about 1.5%. How large of a survey would you recommend?

Sol: (a) sample size $= 331$, $\hat{p} = 0.48$,

$$SE_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.48 \cdot (0.52)}{331}} = 0.0275$$

$$90\% \text{ CI}: \quad \hat{p} \pm z^*_{90\%} \cdot SE_{\hat{p}} = 0.48 \pm 1.65 \cdot (0.0275)$$

$$\rightarrow (0.435, 0.525)$$

To conclude, we have 90% confidence that the 43.5% to 52.5% of all Americans who decide not to go to college due to the cost.

(b) Margin of error $= z^*_{90} \cdot SE_{\hat{p}}$

If Margin of error $= 1.5\% = 0.015$, we have

$$0.015 = 1.65 \cdot \sqrt{\frac{0.48 \cdot (0.52)}{n}} \Rightarrow \left(\frac{0.015}{1.65}\right)^2 = \frac{0.48 \cdot 0.52}{n}$$

$$\Rightarrow n = 0.48 \cdot 0.52 \cdot \left(\frac{1.65}{0.015}\right)^2 = 3020.16 \approx 3021$$

At least 3021 samples.

6.15 National Health Plan, Part II. Exercise 6.11 presents the results of a poll evaluating support for a generic "National Health Plan" in the US in 2019, reporting that 55% of Independents are supportive. If we wanted to estimate this number to within 1% with 90% confidence, what would be an appropriate sample size?

Sol: $n = ?$ $\hat{p} = 55\%$, margin of error $= 1\% = 0.01$

$$0.01 = z^*_{90} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.65 \cdot \sqrt{\frac{0.55(0.45)}{n}}$$

$$\Rightarrow \left(\frac{0.01}{1.65}\right)^2 = \frac{0.55(0.45)}{n} \Rightarrow n = \frac{0.55 \cdot 0.45 \cdot (1.65)^2}{(0.01)^2} = 6738.1 \approx 6739$$

At least 6739 samples.

6.17 Social experiment, Part I. A "social experiment" conducted by a TV program questioned what people do when they see a very obviously bruised woman getting picked on by her boyfriend. On two different occasions at the same restaurant, the same couple was depicted. In one scenario the woman was dressed "provocatively" and in the other scenario the woman was dressed "conservatively". The table below shows how many restaurant diners were present under each scenario, and whether or not they intervened.

		Scenario		
		Provocative	Conservative	Total
Intervene	Yes	5	15	20
	No	15	10	25
	Total	20	25	45

Explain why the sampling distribution of the difference between the proportions of interventions under provocative and conservative scenarios does not follow an approximately normal distribution.

Sol: Since success – failure condition is not met:
there are only 5 interventions under provocative scenario,
then $\hat{p}_p - \hat{p}_c$ is not approximately normal.

6.18 Heart transplant success. The Stanford University Heart Transplant Study was conducted to determine whether an experimental heart transplant program increased lifespan. Each patient entering the program was officially designated a heart transplant candidate, meaning that he was gravely ill and might benefit from a new heart. Patients were randomly assigned into treatment and control groups. Patients in the treatment group received a transplant, and those in the control group did not. The table below displays how many patients survived and died in each group.[22]

	control	treatment
alive	4	24
dead	30	45

Suppose we are interested in estimating the difference in survival rate between the control and treatment groups using a confidence interval. Explain why we cannot construct such an interval using the normal approximation. What might go wrong if we constructed the confidence interval despite this problem?

Sol: Since success–failure condition is not met:
there are only 4 samples alive in control group.
then $\hat{p}_c - \hat{p}_t$ is not approximately normal.

6.19 Gender and color preference. A study asked 1,924 male and 3,666 female undergraduate college students their favorite color. A 95% confidence interval for the difference between the proportions of males and females whose favorite color is black ($p_{male} - p_{female}$) was calculated to be (0.02, 0.06). Based on this information, determine if the following statements about undergraduate college students are true or false, and explain your reasoning for each statement you identify as false.[23]

(a) We are 95% confident that the true proportion of males whose favorite color is black is 2% lower to 6% higher than the true proportion of females whose favorite color is black.

(b) We are 95% confident that the true proportion of males whose favorite color is black is 2% to 6% higher than the true proportion of females whose favorite color is black.

(c) 95% of random samples will produce 95% confidence intervals that include the true difference between the population proportions of males and females whose favorite color is black.

(d) We can conclude that there is a significant difference between the proportions of males and females whose favorite color is black and that the difference between the two sample proportions is too large to plausibly be due to chance.

(e) The 95% confidence interval for ($p_{female} - p_{male}$) cannot be calculated with only the information given in this exercise.

Sol (a) False. It should be males who love color black is 2% to 6% higher than females.

(b) True

(c) True

(d) True

(e) False, we just put a negative on 0.02 and 0.06 and get the interval (−0.06, −0.02)

6.27 Sleep deprived transportation workers. The National Sleep Foundation conducted a survey on the sleep habits of randomly sampled transportation workers and a control sample of non-transportation workers. The results of the survey are shown below.[28]

	Control	Pilots	Truck Drivers	Train Operators	Bus/Taxi/Limo Drivers
		Transportation Professionals			
Less than 6 hours of sleep	35	19	35	29	21
6 to 8 hours of sleep	193	132	117	119	131
More than 8 hours	64	51	51	32	58
Total	292	202	203	180	210

more than 6 hours

Conduct a hypothesis test to evaluate if these data provide evidence of a difference between the proportions of truck drivers and non-transportation workers (the control group) who get less than 6 hours of sleep per day, i.e. are considered sleep deprived.

Sol: Let P_T and P_C be the proportions of truck drivers and control group who get less than 6 hours of sleep, respectively.

$\hat{P}_T = \frac{35}{203} = 0.1724$, $\hat{P}_C = \frac{35}{292} = 0.1199$

Prepare: H_0: $P_T - P_c = 0$ (no difference)

H_A $P_T - P_c \neq 0$ (there is a difference)

$\alpha = 0.05$. null value $P_0 = 0$.

Pool propotion $\hat{P}_{Pool} = \frac{35+35}{203+292} = \frac{70}{495} = 0.1414$

Check: Independence: simple random samples individually.

success-failure conditions: $n_T = 203$, $n_c = 292$

$n_T \times \hat{P}_{Pool} = 203 \times 0.1414 = 28.70$ >10 $n_T \times (1-\hat{P}_{Pool}) = 203 \times (1-0.1414) = 174.3$ >10

$n_c \times \hat{P}_{Pool} = 292 \times 0.1414 = 41.29$ >10 $n_c \times (1-\hat{P}_{Pool}) = 292 \times (1-0.1414) = 250.71$ >10

Since the conditions are satisfied, $\hat{P}_T - \hat{P}_c$ is expected to be approximatly normal

Calculate: $SE_{\hat{P}_{Pool}} = \sqrt{\frac{\hat{P}_{Pool}(1-\hat{P}_{Pool})}{n_T} + \frac{\hat{P}_{Pool}(1-\hat{P}_{Pool})}{n_c}} = \sqrt{\frac{0.1414(1-0.1414)}{203} + \frac{(0.1414)(1-0.1414)}{292}}$

$= 0.031840$

$= 0.0318$

$Z = \frac{(\hat{P}_T - \hat{P}_c) - P_0}{SE_{\hat{P}_{Pool}}} = \frac{0.0525}{0.0318} = 1.65$

P-value $= 2 \times P(|Z| > 1.65) = 2 \cdot 0.0495 = 0.099$

Conclude: Since p-value $> \alpha$, we fail to reject H_0, i.e. there is no difference of P_T and P_C.

6.28 Prenatal vitamins and Autism. Researchers studying the link between prenatal vitamin use and autism surveyed the mothers of a random sample of children aged 24 - 60 months with autism and conducted another separate random sample for children with typical development. The table below shows the number of mothers in each group who did and did not use prenatal vitamins during the three months before pregnancy (periconceptional period).[29]

		Autism		
		Autism	Typical development	Total
Periconceptional	No vitamin	111	70	181
prenatal vitamin	Vitamin	143	159	302
	Total	254	229	483

(a) State appropriate hypotheses to test for independence of use of prenatal vitamins during the three months before pregnancy and autism.

(b) Complete the hypothesis test and state an appropriate conclusion. (Reminder: Verify any necessary conditions for the test.)

(c) A New York Times article reporting on this study was titled "Prenatal Vitamins May Ward Off Autism". Do you find the title of this article to be appropriate? Explain your answer. Additionally, propose an alternative title.[30]

Sol: Let P_V and P_{NV} be the proportions of autism which with and without prenatal vitamin, respectively.

$$\hat{P}_V = \frac{143}{302} = 0.47, \quad \hat{P}_{NV} = \frac{111}{181} = 0.61$$

(a) H_0: $P_V = P_{NV}$ (no difference)

H_A: $P_V \neq P_{NV}$

$\alpha = 0.05$. null value $P_0 = 0$

(b) Check: Independence: Simple random samples

Success-failure: $\hat{P}_{pool} = \frac{143+111}{302+181} = \frac{254}{483} = 0.53$, $n_{nv} = 181$, $n_v = 302$

$n_v \times \hat{P}_{pool} = 302 \times 0.53 = 160.06 > 10$ $\quad n_v \times (1-\hat{P}_{pool}) = 302 \times 0.47 = 141.94 > 10$

$n_{nv} \times \hat{P}_{pool} = 181 \times 0.53 = 95.93 > 10$ $\quad n_{nv} \times (1-\hat{P}_{pool}) = 181 \times 0.47 = 85.07 > 10$

Since both conditions are satisfied, $(\hat{P}_V - \hat{P}_{NV})$ is approximately normal

calculate:

$$SE_{\hat{P}_{pool}} = \sqrt{\frac{\hat{P}_{pool}(1-\hat{P}_{pool})}{n_v} + \frac{\hat{P}_{pool}(1-\hat{P}_{pool})}{n_{nv}}} = \sqrt{\frac{0.53 \cdot 0.47}{302} + \frac{0.53 \cdot 0.47}{181}} = 0.046915$$

$$Z = \frac{(\hat{P}_V - \hat{P}_{nv}) - P_0}{SE_{\hat{P}_{pool}}} = \frac{(-0.14)}{0.0469} = -2.985 \approx -2.99$$

p-value: $2 \cdot P(|Z| > 2{,}99) = 2 \cdot 0{,}0014 = 0{,}0028$

Conclude: Since p-value $< \alpha$, we reject H_0

(c)