30% of adults in the USA believe there should be a wall along the border with Mexico. Describe the sampling distribution of 𝑝̂ from a random sample of 1,000 adults in the USA. The probability of success p is 0.3. μ𝑝̂ = 0.3. σ𝑝̂ = sqrt(0.3\*0.7/1000) = 0.016 The sampling distribution will be approximately normal.

Using your answer above, what is the probability of obtaining the following outcomes from the random sample of 1,000 adults:

* More than 33% believe there should be a wall along the border with Mexico z = (0.33 – 0.3)/0.016 = 0.03/0.016 = 1.875 p = 0.4696
* Less than 28% believe there should be a wall along the border with Mexico z = (0.28 – 0.3)/0.016 = -0.02/0.016 = -1.25 p = 0.3943

**Binomial Exact Test**: Suppose the current treatment for a disease cures 62% of all cases. A new treatment method has been proposed and studied. In a sample of 80 subjects with the disease that were treated with the new method, 63 were cured. Do the results of this study support the claim that the new method has a higher cure rate than the existing method? Conduct a binomial exact test to determine the answer to this question. Make sure to include the null and alternative hypotheses, test statistic, p-value, and conclusion for the hypothesis test. H0: p = 0.62 Ha: p > 0.62 Test statistic: Y = 63 P(X ≥ 63 | p = 0.62) = 0.001035

Conclusion: We have evidence to suggest that the new treatment method has a higher cure rate than the existing treatment.

**Score Test**: A start-up company is about to market a new computer printer. It decides to gamble by purchasing commercials during the Super Bowl. The company is hoping the name recognition will be worth the high cost of the commercials. The company’s goal is to have over 40% of the public recognize its brand name and associate it with computer printers. The day after the game, a pollster contacts 420 randomly selected adults and finds that 181 of them know this company makes printers. Is this evidence that the company met their goal? Use R to conduct a hypothesis test to determine the answer to this question. Make sure to include the null and alternative hypotheses, test statistic, p-value, and conclusion for the hypothesis test. H0: p ≤ 0.40 Ha: p > 0.40 Test statistic: 1.294836 p-value: P(Z > 1.294836) = 0.09769 Conclusion: There is not enough evidence to conclude that the proportion of the public recognizing the company’s brand name and associating it with computer printers is greater than 40%

A start-up company is about to market a new computer printer. It decides to gamble by purchasing commercials during the Super Bowl. The company is hoping the name recognition will be worth the high cost of the commercials. The day after the game, a pollster contacts 420 randomly selected adults and finds that 181 of them know this company makes printers. Find and interpret a 95% CI for the proportion of the population that recognizes this start-up company makes printers. p̂ = 181/420 Plug into CI Normal Approx Method

95% CI for the proportion of the population that recognizes the startup company makes printers is given by the interval (0.383, 0.475). This means we are 95% confident that the true population lies between 38.3% and 47.5%

**A certain genetic mutation occurs in a population with probability 0.05. A researcher has genetic material from 40 unrelated members of this population and tests for the mutation.** *The number of people in a sample of 40 unrelated members of this population with this genetic mutation has a binomial distribution.* n = 40 and p = 0.05. *Calculate the probability that at least 1 person in a sample of 40 unrelated members of this population will have the genetic mutation.* Find P(Y ≥ 1). 1-(0.050(1-0.05)40) = 0.8714878. *Calculate the probability that no more than 3 people in a sample of 40 unrelated members of this population will have the genetic mutation.* Find P(Y ≤ 3). sum(dbinom(0:3, 40, 0.05)) ## [1] 0.8618502. *What is the mean number of people with the genetic mutation in a sample of 40 unrelated members of this population?* E(Y ) = np = 40(0.05) = 2. *What is the variance and standard deviation of the number of people with the genetic mutation in a sample of 40 unrelated members of this population?* V(Y) = np(1 − p) = 40(0.05)(0.95) = 1.9. Standard deviation = √ 1.9 = 1.3784. This distribution is unimodal and skewed right.

**Suppose, based on numerous chess games between these two players, it has been determined the probability Player A would win is 0.40, the probability Player B would win is 0.35, and the probability the game would end in a draw is 0.25.** *Find the probability that Player A would win 7 games, Player B would win 2 games, and the remaining 3 games would each end in a draw if they played 12 games.* Find P(YA = 7, YB = 2, YT = 3). (12!/(7!2!3!))(0.4)7(0.35)2(0.25)3 = 0.02483712 *Find the expected number of games Player A would win and the expected number of games Player B would win if the two players played 12 games.* Player A: E(YA) = npA = 12(0.4) = 4.8 Player B: E(YB) = npB = 12(0.35) = 4.2 *Find the variance of the number of games Player A would win and the variance of the number of games Player B would win if the two players played 12 games.* Player A: V (YA) = npA(1 − pA) = 12(0.4)(0.6) = 2.88 Player B: V (YB) = npB(1 − pB) = 12(0.35)(0.65) = 2.73 *Find the correlation of the number of games won between Player A and Player B*. ρ(YA, YB) = - sqrt((0.4\*0.35)/(0.6\*0.65)) = −0.5991447

**According to the company’s website, the proportion of Green milk chocolate M&Ms produced is 0.16. Let the sample proportion p̂ be the proportion of Green milk chocolate M&Ms in a large bag of 100 of the candies.** *Determine the sampling distribution of the sample proportion p̂.* Since np = 100(0.16) = 16 and n(1 − p) = 100(0.84) = 84 are both larger than 10, the sampling distribution of p̂ is approximately Normal with mean p = 0.16 and standard deviation sqrt((0.16\*0.84)/100) = 0.0367 *Find the probability a large bag of 100 of the candies would have more than 20% green M&Ms.* Find P(p̂ > 0.2). Using the sampling distribution from part (a), this probability is approximately equal to: 0.1376168 [1-pnorm(0.2, 0.16, 0.0367)] *Find the probability a large bag of 100 of the candies would have less than 10% green M&Ms.* Find P(p̂ < 0.1). Using the sampling distribution from part (a), this probability is approximately equal to: 0.05085347 [pnorm(0.1, 0.16, 0.0367)]

**A number of numbers on a white background

Description automatically generatedMany dog owners teach their dogs to “shake hands”. For 15 consecutive times the trick is done in the same way with the same person, you find that the dog extended his right paw 10 times.** If the dog doesn’t favor a paw, the proportion of times the right paw is offered will be p = 0.5. If he does favor right paw, the proportion of times right paw is offered will be p > 0.5. H0: p = 0.5. Ha : p > 0.5 Test Statistic: Y = 10. P-value: P(Y ≥ 10|p = 0.5) = 0.1509. Conclusion: We have little evidence the dog prefers his right paw *Use R to find the rejection region for this test. Use α = 0.05.* We want to find the value of y so that P(Y ≥ y) ≤ 0.05. Using the dbinom() function, we have: sum(dbinom(11:15, 15, 0.5)) ## [1] 0.05923462 sum(dbinom(12:15, 15, 0.5)) ## [1] 0.01757812 sum(dbinom(13:15, 15, 0.5)) ## [1] 0.003692627. This makes the rejection region Y ≥ 12. *Based on the rejection region you found in part (c), what is the observed Type I error rate for this test?* The observed Type I error rate will be P(Y ≥ 12|p = 0.5). This is: sum(dbinom(12:15, 15, 0.5)) ## [1] 0.01757812 *Based on the rejection region you found in part (c), what is the power of your hypothesis test if the dog favors his right paw with probability 0.6, 0.75, or 0.9?* The power is P(Y ≥ 12|p = pa). For the three values of pa above, we have: sum(dbinom(12:15, 15, 0.6)) ## [1] 0.0905019 sum(dbinom(12:15, 15, 0.75)) ## [1] 0.4612869 sum(dbinom(12:15, 15, 0.9)) ## [1] 0.9444444

**A close up of numbers

Description automatically generatedA company’s old antacid formula provided relief from heartburn for 75% of the people who used it. The company develops a new formula in hopes of improving on the proportion of users who obtain relief. In a random sample of 400 people, 312 had relief of their heartburn.** *Explain why you can use the score test for this hypothesis test.* In this situation, the sample size n = 400 and the value of p0 = 0.75 meaning both np0 = 400(0.75) = 300 and n(1−p0) = 400(0.25) = 100 are greater than 10. We can use the score test since the Normal distribution is a good approximation for the sampling distribution of p̂. *Conduct a score test to determining whether the new formula is better than the old formula.* If the new formula is the same as the old one, the proportion of people who get relief will be p = 0.75. If the new formula is better than the old one, the proportion of people who get relief will be p > 0.75. H0 : p = 0.75. Ha : p > 0.75. Test Statistic: √ X2 = √ 1.92 = (0.78-0.75)/sqrt((0.75\*0.25)/400) = 1.3856. P-value: P(Z > 1.3856) = 0.0829. We have weak evidence the new formula is better than the old one. *Discuss the effect of the value of the population proportion p and the value of α on the power of this hypothesis test.* When pa is further away from p0 = 0.75, the power increases with the power for both pa = 0.85 and pa = 0.9 near 1. When α is larger, the power is also larger. *After the above analysis, the company decided to switch production to the new antacid formula. After several years in production, they found the new formula provided relief to 80% of the people who used it. Suppose the company would like to test another formula in the future. What sample size will they need to use to have a power of 0.9 to detect an improvement in the proportion of users who obtain relief of 0.05 if α = 0.05.* npowerprop.test(0.8, 0.85, alternative = "greater", 0.05, 0.9) ## [1] 498

*A number on a white background

Description automatically generated****A USA Today/Gallup poll asked 1,006 randomly selected people 18 years old and older in telephone interviews “Suppose you checked into a hotel and were given a room on the thirteenth floor. Would this bother you or not?”*** *Our category of interest is being bothered by staying on the 13th floor. Calculate a 95% CI for the population proportion p using the normal approximation method.* The 95% CI is (0.1094, 0.1510). We are 95% confident the proportion of adults in the US would be bothered staying on 13th floor is between 0.1094 and 0.1510. *Calculate a 95% CI for the population proportion p using Wilson’s score method.* The 95% CI is (0.1108, 0.1524). *If the Poll were conducted again using this question, what sample size would be needed to guarantee a 90% CI would have a margin of error of no more than 2%?* In order to guarantee the 90% CI will have a margin of error of no more than 2%, we will need to use the worst case scenario formula and use p = 0.5. So the sample size should be 1,691.

**According to a Gallup poll, of 1,019 randomly selected adults aged 18 or older in the United States, 662 believe that global warming is more a result of human actions than natural causes.** *Describe the population proportion of interest p in words.* The population proportion of interest is the proportion of adults in the United States who believe that global warming is more a result of human actions than natural causes. *Give the value of the sample proportion p̂.* p̂ = 662/1019 = 0.6497 *Calculate a 95% CI for the population proportion of interest using Wilson’s score method.* The 95% CI is (0.6199, 0.6783). We are 95% confident the proportion of adults in the United States who believe that global warming is more a result of human actions than natural causes is between 0.6199 and 0.6783. *Gallup is planning to conduct another poll on global warming. They would like to have a 95% confidence interval with a margin of error of no more than 2.5%. What sample size do they need to obtain this margin of error?* In order for the 95% confidence interval to have a margin of error of no more than 2.5%, we will need to use the worst case scenario formula and use p = 0.5. So the sample size should be 1,537

A screenshot of a computer

Description automatically generated**12 zodiac signs based on their birthday. In one small study, Fortune magazine collected the zodiac signs of 265 heads of the largest 400 companies. Does it appear that some zodiac signs are more likely to be represented in heads of these types of companies than others?** *Give the null and alternative hypotheses for answering the question posed. Assume each zodiac sign covers an equal number of birthdays in the year.* If there is no association between zodiac sign and being the head of this type of company, the probability of each zodiac sign is the same = 1/12. H0: p1 = p2 = … = p12 = 1/12. Ha: at least one pj ≠ 1/12 *Calculate the expected number of births in each zodiac sign under the null hypothesis.* The expected value is E(Yj ) = npj = 265(1/12) = 22.083 *Calculate the contribution of the category Scorpio (8th in list) to the test statistic X2 . Only calculate this value for the category Scorpio, not for all the other categories.* Calculate using the expected value above and the summary table information. This is: (21-22.0833)2 / 22.0833 = 0.0531. *Determine the value of the test statistic X2 for this hypothesis test.* From the output of the goodness of fit test, the test statistic X2 = 7.1962. *What is the number of degrees of freedom for the test statistic X2 ?* There are J = 12 categories, so the degrees of freedom is J − 1 = 11. *Determine the p-value for this hypothesis test.* From the output of the goodness of fit test, the p-value is 0.783. There is no evidence of lack of model fit. Some zodiac signs are no more likely to be represented in heads of these types of companies than others.

**90% z = 1.645**

**95% z = 1.96**