Review of Part V – From the Data at Hand to the World at Large

1. Herbal cancer.

H_0: The cancer rate for those taking the herb is the same as the cancer rate for those not taking the herb. $(p_{Herb} = p_{Not}$ or $p_{Herb} - p_{Not} = 0)$

H_A: The cancer rate for those taking the herb is higher than the cancer rate for those not taking the herb. $(p_{Herb} > p_{Not}$ or $p_{Herb} - p_{Not} > 0)$

2. Colorblind.

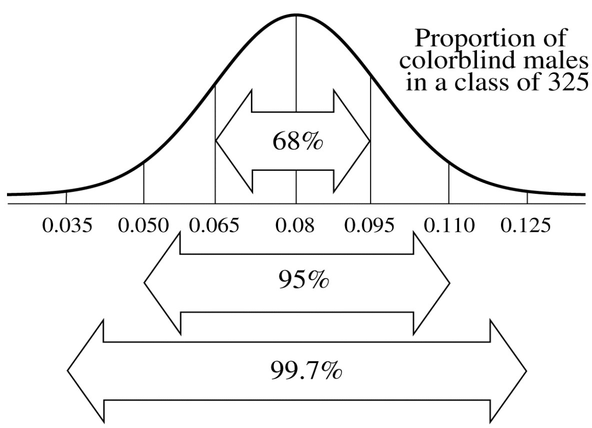
a) **Randomization condition:** The 325 male students are probably representative of all males.
10% condition: 325 male students are less than 10% of the population of males.
Success/Failure condition: $np = (325)(0.08) = 26$ and $nq = (325)(0.92) = 299$ are both greater than 10, so the sample is large enough.

Since the conditions have been satisfied, a Normal model can be used to model the sampling distribution of the proportion of colorblind men among 325 students.

b) $\mu_{\hat{p}} = p = 0.08$

$$\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.08)(0.92)}{325}} \approx 0.015$$

c)



d) According to the Normal model, we expect about 68% of classes with 325 males to have between 6.5% and 9.5% colorblind males. We expect about 95% of such classes to have between 5% and 11% colorblind males. About 99.7% of such classes are expected to have between 3.5% and 12.5% colorblind males.

3. Birth days.

a) If births are distributed uniformly across all days, we expect the number of births on each day to be $np = (72)(\frac{1}{7}) \approx 10.29$.

b) **Randomization condition:** The 72 births are likely to be representative of all births at the hospital with regards to day of birth.
10% condition: 72 births are less than 10% of the births.
Success/Failure condition: The expected number of births on a particular day of the week is $np = (72)(\frac{1}{7}) \approx 10.29$ and the expected number of births not on that particular day is $nq = (72)(\frac{6}{7}) \approx 61.71$. These are both greater than 10, so the sample is large enough.

Since the conditions have been satisfied, a Normal model can be used to model the sampling distribution of the proportion of 72 births that occur on a given day of the week.

$\mu_{\hat{p}} = p = 1/7 \approx 0.1429$

$\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(1/7)(6/7)}{72}} \approx 0.04124$

There were 7 births on Mondays, so $\hat{p} = \frac{7}{72} \approx 0.09722$. This is only about a 1.11 standard deviations below the expected proportion, so there's no evidence that this is unusual.

c) The 17 births on Tuesdays represent an unusual occurrence. For Tuesdays, $\hat{p} = \frac{17}{72} \approx 0.2361$, which is about 2.26 standard deviations above the expected proportion of births. There is evidence to suggest that the proportion of births on Tuesdays is higher than expected, if births are distributed uniformly across days.

d) Some births are scheduled for the convenience of the doctor and/or the mother.

4. Polling 2004.

a) No, the number of votes would not always be the same. We expect a certain amount of variability when sampling.

b) This is NOT a problem about confidence intervals. We already know the true proportion of voters who voted for Bush. This problem deals with the sampling distribution of that proportion.

We would expect 95% of our sample proportions of Bush voters to be within 1.960 standard deviations of the true proportion of Bush voters, 50.7%.

$\sigma(\hat{p}_B) = \sqrt{\frac{p_B q_B}{n}} = \sqrt{\frac{(0.507)(0.493)}{1000}} \approx 1.58\%$

So, we expect 95% of our sample proportions to be within 1.960(1.58%) = 3.1% of 48.3%, or between 47.6% and 53.8%.

c) Since we only expect 0.004(1000) = 4 votes for Ralph Nader, we cannot represent the sampling model with a Normal model. The Success/Failure condition is not met.

d) The sample proportion of Nader voters is expected to vary less than the sample proportion of Bush voters. Proportions farther away from 50% have smaller standard errors. (Look at the standard deviations calculated for parts b and c.)

5. Leaky gas tanks.

a) H_0: The proportion of leaky gas tanks is 40%. ($p = 0.40$)
H_A: The proportion of leaky gas tanks is less than 40%. ($p < 0.40$)

b) **Randomization condition:** A random sample of 27 service stations in California was taken.
10% condition: 27 service stations are less than 10% of all service stations in California.
Success/Failure condition: $np = (27)(0.40) = 10.8$ and $nq = (27)(0.60) = 16.2$ are both greater than 10, so the sample is large enough.

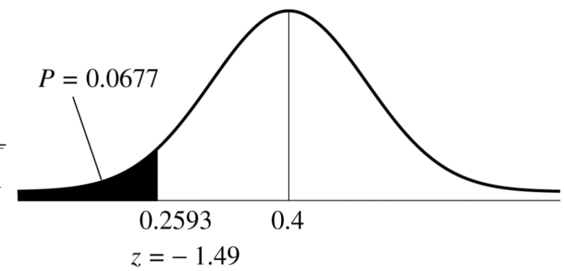
c) Since the conditions have been satisfied, a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p = 0.40$ and $\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.40)(0.60)}{27}} \approx 0.09428$. We can perform a one-proportion z-test.

The observed proportion of leaky gas tanks is $\hat{p} = \frac{7}{27} \approx 0.2593$.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.2593 - 0.40}{\sqrt{\frac{(0.40)(0.60)}{27}}}$$

$$z \approx -1.49$$



d) Since the P-value = 0.0677 is relatively high, we fail to reject the null hypothesis. There is little evidence that the proportion of leaky gas tanks is less than 40%. The new program doesn't appear to be effective in decreasing the proportion of leaky gas tanks.

e) If the program actually works, we haven't done anything *wrong*. Our methods are correct. Statistically speaking, we have committed a Type II error.

f) In order to decrease the probability of making this type of error, we could lower our standards of proof, by raising the level of significance. This will increase the power of the test to detect a decrease in the proportion of leaky gas tanks. Another way to decrease the probability that we make a Type II error is to sample more service stations. This will decrease the variation in the sample proportion, making our results more reliable.

g) Increasing the level of significance is advantageous, since it decreases the probability of making a Type II error, and increases the power of the test. However, it also increases the probability that a Type I error is made, in this case, thinking that the program is effective when it really is not effective.
Increasing the sample size decreases the probability of making a Type II error and increases power, but can be costly and time-consuming.

6. Surgery and germs.

a) Lister imposed a treatment, the use of carbolic acid as a disinfectant. This is an experiment.

b) H_0: The survival rate when carbolic acid is used is the same as the survival rate when carbolic acid is not used. $(p_C = p_N$ or $p_C - p_N = 0)$

H_A: The survival rate when carbolic acid is used is greater than the survival rate when carbolic acid is not used. $(p_C > p_N$ or $p_C - p_N > 0)$

Randomization condition: There is no mention of random assignment. Assume that the two groups of patients were similar, and amputations took place under similar conditions, with the use of carbolic acid being the only variable.
10% condition: 40 and 35 are both less than 10% of all possible amputations.
Independent samples condition: It is reasonable to think that the groups were not related in any way.
Success/Failure condition: $n\hat{p}$(carbolic acid) = 34, $n\hat{q}$(carbolic acid) = 6, $n\hat{p}$(none) = 19, and $n\hat{q}$(none) = 16. The number of patients who died in the carbolic acid group is only 6, but the expected number of deaths using the pooled proportion, $n\hat{q}_{\text{pooled}} = (40)\left(\frac{22}{75}\right) = 11.7$, so the samples are both large enough.

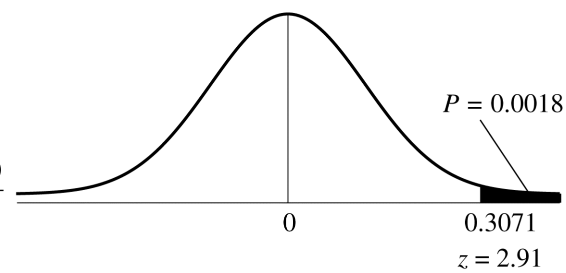
Since the conditions have been satisfied, we will perform a two-proportion z-test. We will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by:

$$SE_{\text{pooled}}\left(\hat{p}_C - \hat{p}_N\right) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_C} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_N}} = \sqrt{\frac{\left(\frac{53}{75}\right)\left(\frac{22}{75}\right)}{40} + \frac{\left(\frac{53}{75}\right)\left(\frac{22}{75}\right)}{35}} \approx 0.1054.$$

The observed difference between the proportions is:
0.85 – 0.5429 = 0.3071.

Since the P-value = 0.0018 is low, we reject the null hypothesis. There is strong evidence that the survival rate is higher when carbolic acid is used to disinfect the operating room than when carbolic acid is not used.

$$z = \frac{0.3071 - 0}{0.1054}$$
$$z \approx 2.91$$



c) We don't know whether or not patients were randomly assigned to treatments, and we don't know whether or not blinding was used.

7. Scrabble.

a) The researcher believes that the true proportion of As is within 10% of the estimated 54%, namely, between 44% and 64%.

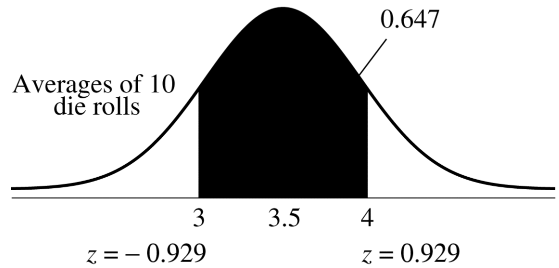
b) A large margin of error is usually associated with a small sample, but the sample consisted of "many" hands. The margin of error is large because the standard error of the sample is large. This occurs because the true proportion of As in a hand is close to 50%, the most difficult proportion to predict.

c) This provides no evidence that the simulation is faulty. The true proportion of As is contained in the confidence interval. The researcher's results are consistent with 63% As.

8. Dice.

Die rolls are truly independent, and the distribution of the outcomes of die rolls is not skewed (it's uniform). According to the CLT, the sampling distribution of $\bar{y}$, the average for 10 die rolls, can be approximated by a Normal model, with $\mu_{\bar{y}} = 3.5$ and standard deviation $\sigma(\bar{y}) = \dfrac{1.7}{\sqrt{10}} \approx 0.538$, even though 10 rolls is a fairly small sample.

According to the Normal model, the probability that the average of 10 die rolls is between 3 and 4 (and therefore the probability of the sum of 10 die rolls is between 30 and 40) is approximately 0.647.



9. News sources.

a) The Pew Research Foundation believes that the true proportion of people who obtain news from the Internet is between 30% and 36%.

b) The smaller sample size in the limited sample would result in a larger standard error. This would make the margin of error larger, as well.

c) $\hat{p} \pm z^* \sqrt{\dfrac{\hat{p}\hat{q}}{n}} = (0.45) \pm 1.960\sqrt{\dfrac{(0.45)(0.55)}{239}} = (38.7\%, 51.3\%)$

We are 95% confident that between 38.7% and 51.3% of active traders rely on the Internet for investment information.

d) The sample of 239 active traders is smaller than either of the earlier samples. This results in a larger margin of error.

10. Death penalty 2006.

a) **Independence assumption:** There is no reason to believe that one randomly selected American adult's response will affect another's.
Randomization condition: Gallup randomly selected 537 American adults.
10% condition: 537 results is less than 10% of all American adults.
Success/Failure condition: $n\hat{p} = (537)(0.47) = 252$ and $n\hat{q} = (537)(0.53) = 285$ are both greater than 10, so the sample is large enough.

Since the conditions are met, we can use a one-proportion z-interval to estimate the percentage of American adults who favor the death penalty.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(0.47\right) \pm 1.960\sqrt{\frac{\left(0.47\right)\left(0.53\right)}{537}} = (42.7\%, 51.1\%)$$

We are 95% confident that between 42.7% and 51.1% of Americans favor the death penalty.

b) Since the interval extends above 50%, it is plausible that the death penalty still has majority support.

c)

$$ME = z^*\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.02 = 2.326\sqrt{\frac{(0.50)(0.50)}{n}}$$

$$n = \frac{(2.326)^2(0.50)(0.50)}{(0.02)^2}$$

$$n \approx 3382 \text{ people}$$

We do not know the true proportion of American adults in favor of the death penalty, so use $\hat{p} = \hat{q} = 0.50$, for the most cautious estimate. In order to determine the proportion of American adults in favor of the death penalty to within 2% with 98% confidence, we would have to sample at least 3382 people.

11. Bimodal.

a) The *sample's* distribution (NOT the *sampling* distribution), is expected to look more and more like the distribution of the population, in this case, bimodal.

b) The expected value of the sample's mean is expected to be μ, the population mean, regardless of sample size.

c) The variability of the sample mean, $\sigma(\bar{y})$, is $\frac{\sigma}{\sqrt{n}}$, the population standard deviation divided by the square root of the sample size, regardless of the sample size.

d) As the sample size increases, the sampling distribution model becomes closer and closer to a Normal model.

12. Vitamin D.

a) Certainly, the 1546 women are less than 10% of all African-American women, and $n\hat{p}$ = (1546)(0.42) = 649 and $n\hat{q}$ = (1546)(0.58) = 897 are both greater than 10, so the sample is large enough. We would like to know that the sample is random. This would help assure us that these women were chosen independently.

b) $\hat{p} \pm z^*\sqrt{\frac{\hat{p}\hat{q}}{n}} = (0.42) \pm 1.960\sqrt{\frac{(0.42)(0.58)}{1546}} = (39.5\%, 44.5\%)$.

c) We are 95% confident that between 39.5% and 44.5% of African-American women have a vitamin D deficiency.

d) 95% of all random samples of this size will produce intervals that contain the true proportion of African-American women who have a vitamin D deficiency.

13. Archery.

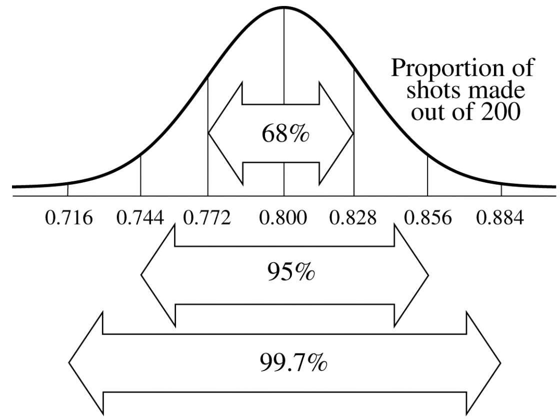
a) $\mu_{\hat{p}} = p = 0.80$

$$\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.80)(0.20)}{200}} \approx 0.028$$

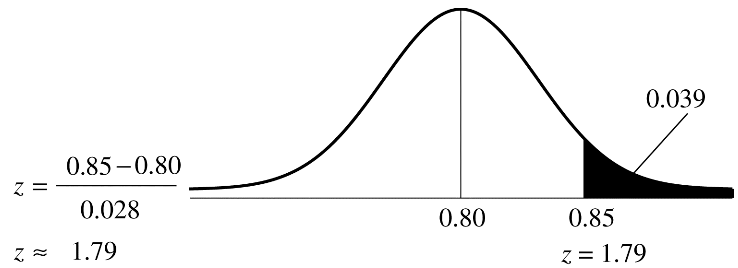
b) np = (200)(0.80) = 160 and nq = (200)(0.20) = 40 are both greater than 10, so the Normal model is appropriate.

c) The Normal model of the sampling distribution of the proportion of bull's-eyes she makes out of 200 is at the right.

Approximately 68% of the time, we expect her to hit the bull's-eye on between 77.2% and 82.8% of her shots. Approximately 95% of the time, we expect her to hit the bull's-eye on between 74.4% and 85.6% of her shots. Approximately 99.7% of the time, we expect her to hit the bull's-eye on between 71.6% and 88.4% of her shots.



d) According to the Normal model, the probability that she hits the bull's-eye in at least 85% of her 200 shots is approximately 0.039.



14. Free throws 2007.

a) **Randomization condition:** Assume that these free throws are representative of the free throw ability of these players.
10% condition: 209 and 208 are less than 10% of all possible free throws.
Independent samples condition: The free throw abilities of these two players should be independent.
Success/Failure condition: $n\hat{p}$(Korver) = 191, $n\hat{q}$(Korver) = 18, $n\hat{p}$(Carroll) = 188, and $n\hat{q}$(Carroll) = 20 are all greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will find a two-proportion z-interval.

$$\left(\hat{p}_K - \hat{p}_C\right) \pm z^* \sqrt{\frac{\hat{p}_K\hat{q}_K}{n_K} + \frac{\hat{p}_C\hat{q}_C}{n_C}} = \left(\tfrac{191}{209} - \tfrac{188}{208}\right) \pm 1.960\sqrt{\frac{\left(\frac{191}{209}\right)\left(\frac{19}{209}\right)}{209} + \frac{\left(\frac{188}{208}\right)\left(\frac{20}{208}\right)}{208}} = \left(-0.045, 0.065\right)$$

We are 95% confident that Kyle Korver's true free throw percentage is between 4.5% worse and 6.5% better than Matt Carroll's.

b) Since the interval for the difference in percentage of free throws made includes 0, it is uncertain who is the better free throw shooter.

15. Twins.

H0: The proportion of preterm twin births in 1990 is the same as the proportion of preterm twin births in 2000. $\left(p_{1990} = p_{2000} \text{ or } p_{1990} - p_{2000} = 0\right)$

HA: The proportion of preterm twin births in 1990 is the less than the proportion of preterm twin births in 2000. $\left(p_{1990} < p_{2000} \text{ or } p_{1990} - p_{2000} < 0\right)$

Randomization condition: Assume that these births are representative of all twin births.
10% condition: 43 and 48 are both less than 10% of all twin births.
Independent samples condition: The samples are from different years, so they are unlikely to be related.
Success/Failure condition: $n\hat{p}(1990) = 20$, $n\hat{q}(1990) = 23$, $n\hat{p}(2000) = 26$, and $n\hat{q}(2000) = 22$ are all greater than 10, so both samples are large enough.

Since the conditions have been satisfied, we will perform a two-proportion z-test. We will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by:

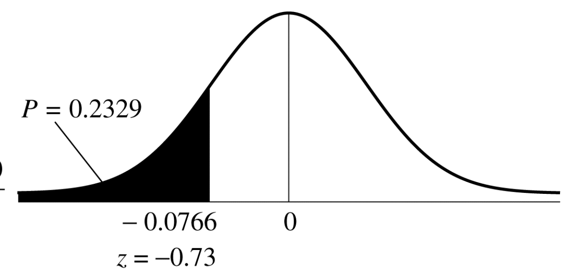
$$SE_{\text{pooled}}\left(\hat{p}_{1990} - \hat{p}_{2000}\right) = \sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{1900}} + \frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{2000}}} = \sqrt{\frac{\left(\frac{46}{91}\right)\left(\frac{45}{91}\right)}{43} + \frac{\left(\frac{46}{91}\right)\left(\frac{45}{91}\right)}{48}} \approx 0.1050.$$

The observed difference between the proportions is:
0.4651 – 0.5417 = – 0.0766

Since the P-value = 0.2329 is high, we fail to reject the null hypothesis. There is no evidence of an increase in the proportion of preterm twin births from 1990 to 2000, at least not at this large city hospital.

$$z = \frac{-0.0766 - 0}{0.1050}$$

$$z \approx -0.73$$



16. Eclampsia.

a) **Randomization condition:** Although not specifically stated, these results are from a large-scale experiment, which was undoubtedly properly randomized.
10% condition: 4999 and 4993 are less than 10% of all pregnant women.
Independent samples condition: Subjects were randomly assigned to the treatments.
Success/Failure condition: $n\hat{p}(\text{mag. sulf.}) = 1201$, $n\hat{q}(\text{mag. sulf.}) = 3798$, $n\hat{p}(\text{placebo}) = 228$, and $n\hat{q}(\text{placebo}) = 4765$ are all greater than 10, so both samples are large enough.

Since the conditions have been satisfied, we will find a two-proportion z-interval.

$$\left(\hat{p}_{MS} - \hat{p}_N\right) \pm z^* \sqrt{\frac{\hat{p}_{MS}\hat{q}_{MS}}{n_{MS}} + \frac{\hat{p}_N\hat{q}_N}{n_N}} = \left(\tfrac{1201}{4999} - \tfrac{228}{4993}\right) \pm 1.960\sqrt{\frac{\left(\frac{1201}{4999}\right)\left(\frac{3798}{4999}\right)}{4999} + \frac{\left(\frac{228}{4993}\right)\left(\frac{4765}{4993}\right)}{4993}} = (0.181, 0.208)$$

We are 95% confident that the proportion of pregnant women who will experience side effects while taking magnesium sulfide will be between 18.1% and 20.8% higher than the proportion of women that will experience side effects while not taking magnesium sulfide.

b) H_0: The proportion of pregnant women who will develop eclampsia is the same for women taking magnesium sulfide as it is for women not taking magnesium sulfide.
$\left(p_{MS} = p_N \text{ or } p_{MS} - p_N = 0\right)$

H_A: The proportion of pregnant women who will develop eclampsia is lower for women taking magnesium sulfide than for women not taking magnesium sulfide.
$\left(p_{MS} < p_N \text{ or } p_{MS} - p_N < 0\right)$

Success/Failure condition: $n\hat{p}$(mag. sulf.) = 40, $n\hat{q}$(mag. sulf.) = 4959, $n\hat{p}$(placebo) = 96, and $n\hat{q}$(placebo) = 4897 are all greater than 10, so both samples are large enough.

Since the conditions have been satisfied (some in part a), we will perform a two-proportion z-test. We will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by:

$$SE_{\text{pooled}}\left(\hat{p}_{MS}-\hat{p}_N\right)=\sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{MS}}+\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_N}}=\sqrt{\frac{\left(\frac{136}{9992}\right)\left(\frac{9856}{9992}\right)}{4999}+\frac{\left(\frac{136}{9992}\right)\left(\frac{9856}{9992}\right)}{4993}}\approx 0.002318.$$

The observed difference between the proportions is 0.00800 – 0.01923 = – 0.01123, which is approximately 4.84 standard errors below the expected difference in proportion of 0.

Since the P-value = 6.4×10^{-7} is very low, we reject the null hypothesis. There is strong evidence that the proportion of pregnant women who develop eclampsia will be lower for women taking magnesium sulfide than for those not taking magnesium sulfide.

17. Eclampsia.

a) H_0: The proportion of pregnant women who die after developing eclampsia is the same for women taking magnesium sulfide as it is for women not taking magnesium sulfide. $\left(p_{MS}=p_N \text{ or } p_{MS}-p_N=0\right)$

H_A: The proportion of pregnant women who die after developing eclampsia is lower for women taking magnesium sulfide than for women not taking magnesium sulfide. $\left(p_{MS}<p_N \text{ or } p_{MS}-p_N<0\right)$

b) **Randomization condition:** Although not specifically stated, these results are from a large-scale experiment, which was undoubtedly properly randomized.
10% condition: 40 and 96 are less than 10% of all pregnant women.
Independent samples condition: Subjects were randomly assigned to the treatments.
Success/Failure condition: $n\hat{p}$(mag. sulf.) = 11, $n\hat{q}$(mag. sulf.) = 29, $n\hat{p}$(placebo) = 20, and $n\hat{q}$(placebo) = 76 are all greater than 10, so both samples are large enough.

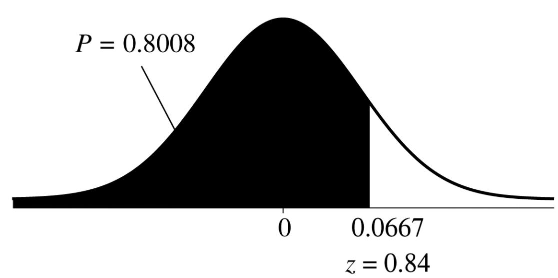
Since the conditions have been satisfied, we will perform a two-proportion z-test. We will model the sampling distribution of the difference in proportion with a Normal model with mean 0 and standard deviation estimated by:

$$SE_{\text{pooled}}\left(\hat{p}_{MS}-\hat{p}_N\right)=\sqrt{\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_{MS}}+\frac{\hat{p}_{\text{pooled}}\hat{q}_{\text{pooled}}}{n_N}}=\sqrt{\frac{\left(\frac{31}{136}\right)\left(\frac{105}{136}\right)}{40}+\frac{\left(\frac{31}{136}\right)\left(\frac{105}{136}\right)}{96}}\approx 0.07895.$$

c) The observed difference between the proportions is:
0.275 – 0.2083 = 0.0667

Since the P-value = 0.8008 is high, we fail to reject the null hypothesis. There is no evidence that the proportion of women who may die after developing eclampsia is lower for women taking magnesium sulfide than for women who are not taking the drug.

$$z=\frac{0.0667-0}{0.07895}$$
$$z\approx 0.84$$



d) There is not sufficient evidence to conclude that magnesium sulfide is effective in preventing death when eclampsia develops.

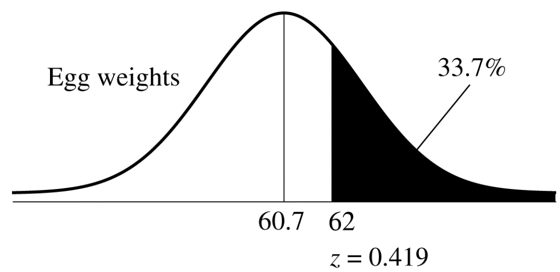
e) If magnesium sulfide is effective in preventing death when eclampsia develops, then we have made a Type II error.

f) To increase the power of the test to detect a decrease in death rate due to magnesium sulfide, we could increase the sample size or increase the level of significance.

g) Increasing the sample size lowers variation in the sampling distribution, but may be costly. The sample size is already quite large. Increasing the level of significance increases power by increasing the likelihood of rejecting the null hypothesis, but increases the chance of making a Type I error, namely thinking that magnesium sulfide is effective when it is not.

18. Eggs.

a) According to the Normal model, approximately 33.7% of these eggs weigh more than 62 grams.



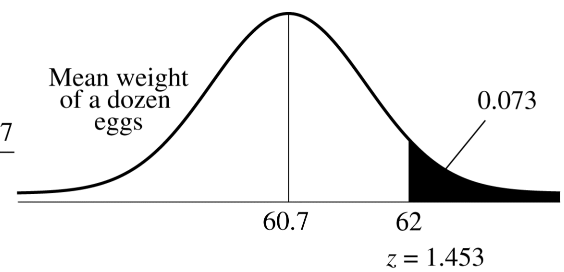
b) **Randomization condition:** The dozen eggs are selected randomly.
10% condition: The dozen eggs are less than 10% of all eggs.

The mean egg weight is $\mu = 60.7$ grams, with standard deviation $\sigma = 3.1$ grams. Since the distribution of egg weights is Normal, we can model the sampling distribution of the mean egg weight of a dozen eggs with a Normal model, with $\mu_{\bar{y}} = 60.7$ grams and standard deviation $\sigma(\bar{y}) = \frac{3.1}{\sqrt{12}} \approx 0.895$ grams.

According to the Normal model, the probability that a randomly selected dozen eggs have a mean greater than 62 grams is approximately 0.073.

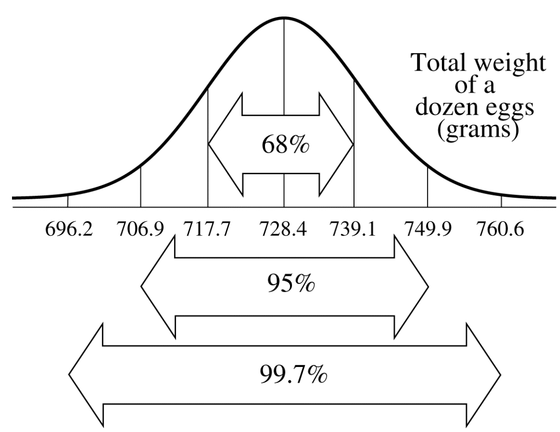
$$z = \frac{62 - 60.7}{\frac{3.1}{\sqrt{12}}}$$

$z \approx 1.453$



c) The average weight of a dozen eggs can be modeled by $N(60.7, 0.895)$, so the total weight of a dozen eggs can be modeled by $N(728.4, 10.74)$.

Approximately 68% of the cartons of a dozen eggs would weigh between 717.1 and 739.1 grams. Approximately 95% of the cartons would weigh between 706.9 and 749.9 grams. Approximately 99.7% of the cartons would weigh between 696.2 and 760.6 grams.



19. Polling disclaimer.

a) It is not clear what specific question the pollster asked. Otherwise, they did a great job of identifying the W's.

b) A sample that was stratified by age, sex, region, and education was used.

c) The margin of error was 4%.

d) Since "no more than 1 time in 20 should chance variations in the sample cause the results to vary by more than 4 percentage points", the confidence level is 19/20 = 95%.

e) The subgroups had smaller sample sizes than the larger group. The standard errors in these subgroups were larger as a result, and this caused the margins of error to be larger.

f) They cautioned readers about response bias due to wording and order of the questions.

20. Enough eggs?

$$ME = z^*\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.02 = 1.960\sqrt{\frac{(0.75)(0.25)}{n}}$$

$$n = \frac{(1.960)^2(0.75)(0.25)}{(0.02)^2}$$

$$n \approx 1801 \text{ eggs}$$

ISA Babcock needs to collect data on about 1800 hens in order to advertise the production rate for the B300 Layer with 95% confidence with a margin of error of $\pm 2\%$.

21. Teen deaths.

a) H_0 : The percentage of fatal accidents involving teenage girls is 14.3%, the same as the overall percentage of fatal accidents involving teens . ($p = 0.143$)
H_A : The percentage of fatal accidents involving teenage girls is lower than 14.3%, the overall percentage of fatal accidents involving teens . ($p < 0.143$)

Independence assumption: It is reasonable to think that accidents occur independently.
Randomization condition: Assume that the 388 accidents observed are representative of all accidents.
10% condition: The sample of 388 accidents is less than 10% of all accidents.
Success/Failure condition: np = (388)(0.143) = 55.484 and nq = (388)(0.857) = 332.516 are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p = 0.143$ and

$$\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.143)(0.857)}{388}} \approx 0.01777.$$

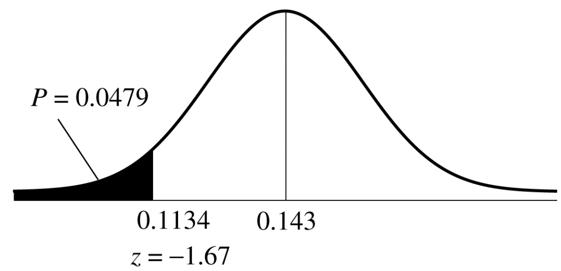
We can perform a one-proportion z-test. The observed proportion of fatal accidents involving teen girls is $\hat{p} = \frac{44}{388} \approx 0.1134$.

Since the P-value = 0.0479 is low, we reject the null hypothesis. There is some evidence that the proportion of fatal accidents involving teen girls is less than the overall proportion of fatal accidents involving teens.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.1134 - 0.143}{\sqrt{\frac{(0.143)(0.857)}{388}}}$$

$$z \approx -1.67$$



b) If the proportion of fatal accidents involving teenage girls is really 14.3%, we expect to see the observed proportion, 11.34%, in about 4.79% of samples of size 388 simply due to sampling variation.

22. Perfect pitch.

a) H_0: The proportion of Asian students with perfect pitch is the same as the proportion of non-Asians with perfect pitch. $\left(p_A = p_N \text{ or } p_A - p_N = 0\right)$

H_A: The proportion of Asian students with perfect pitch is the different than the proportion of non-Asians with perfect pitch. $\left(p_A \neq p_N \text{ or } p_A - p_N \neq 0\right)$

b) Since $P < 0.0001$, which is very low, we reject the null hypothesis. There is strong evidence of a difference in the proportion of Asians with perfect pitch and the proportion of non-Asians with perfect pitch. There is evidence that Asians are more likely to have perfect pitch.

c) If there is no difference in the proportion of students with perfect pitch, we would expect the observed difference of 25% to be seen simply due to sampling variation in less than 1 out of every 10,000 samples of 2700 students.

d) The data do not prove anything about genetic differences causing differences in perfect pitch. Asians are merely more likely to have perfect pitch. There may be lurking variables other than genetics that are causing the higher rate of perfect pitch.

23. Largemouth bass.

a) One would expect many small fish, with a few large fish.

b) We cannot determine the probability that a largemouth bass caught from the lake weighs over 3 pounds because we don't know the exact shape of the distribution. We know that it is NOT Normal.

c) It would be quite risky to attempt to determine whether or not the mean weight of 5 fish was over 3 pounds. With a skewed distribution, a sample of size 5 is not large enough for the Central Limit Theorem to guarantee that a Normal model is appropriate to describe the distribution of the mean.

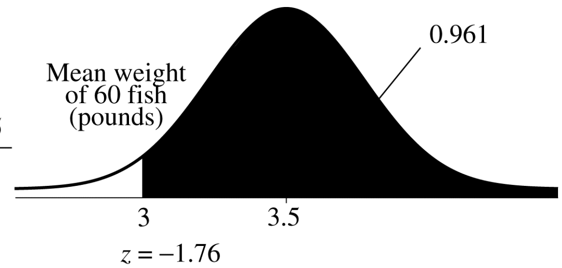
d) A sample of 60 randomly selected fish is large enough for the Central Limit Theorem to guarantee that a Normal model is appropriate to describe the sampling distribution of the mean, as long as 60 fish is less than 10% of the population of all the fish in the lake.

The mean weight is $\mu = 3.5$ pounds, with standard deviation $\sigma = 2.2$ pounds. Since the sample size is sufficiently large, we can model the sampling distribution of the mean weight of 60 fish with a Normal model, with $\mu_{\bar{y}} = 3.5$ pounds and standard deviation $\sigma(\bar{y}) = \dfrac{2.2}{\sqrt{60}} \approx 0.284$ pounds.

According to the Normal model, the probability that 60 randomly selected fish average more than 3 pounds is approximately 0.961.

$$z = \frac{3 - 3.5}{\frac{2.2}{\sqrt{60}}}$$

$$z \approx -1.76$$



24. Cheating.

a) **Independence assumption:** There is no reason to believe that students selected at random would influence each others responses.
Randomization condition: The 4500 students were selected randomly.
10% condition: 4500 students is less than 10% of all students.
Success/Failure condition: $n\hat{p} = (4500)(0.74) = 3330$ and $n\hat{q} = (4500)(0.26) = 1170$ are both greater than 10, so the sample is large enough.

Since the conditions are met, we can use a one-proportion z-interval to estimate the percentage of students who have cheated at least once.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = (0.74) \pm 1.645\sqrt{\frac{(0.74)(0.26)}{4500}} = (72.9\%, 75.1\%)$$

b) We are 90% confident that between 72.9% and 75.1% of high school students have cheated at least once.

c) About 90% of random samples of size 4500 will produce intervals that contain the true proportion of high school students who have cheated at least once.

d) A 95% confidence interval would be wider. Greater confidence requires a larger margin of error.

25. Language.

a) **Randomization condition:** 60 people were selected at random.
10% condition: The 60 people represent less than 10% of all people.
Success/Failure condition: $np = (60)(0.80) = 48$ and $nq = (60)(0.20) = 12$ are both greater than 10.

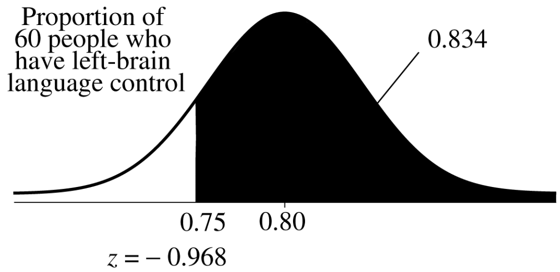
Therefore, the sampling distribution model for the proportion of 60 randomly selected people who have left-brain language control is Normal, with $\mu_{\hat{p}} = p = 0.80$ and standard deviation $\sigma(\hat{p}) = \sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{(0.80)(0.20)}{60}} \approx 0.0516$.

b) According to the Normal model, the probability that over 75% of these 60 people have left-brain language control is approximately 0.834.

$$z = \frac{\hat{p} - \mu_{\hat{p}}}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.75 - 0.80}{\sqrt{\frac{(0.80)(0.20)}{60}}}$$

$$z \approx -0.968$$



c) If the sample had consisted of 100 people, the probability would have been higher. A larger sample results in a smaller standard deviation for the sample proportion.

d) Answers may vary. Let's consider three standard deviations below the expected proportion to be "almost certain". It would take a sample of (exactly!) 576 people to make sure that 75% would be 3 standard deviations below the expected percentage of people with left-brain language control.

Using round numbers for n instead of z, about 500 people in the sample would make the probability of choosing a sample with at least 75% of the people having left-brain language control is a whopping 0.997. It all depends on what "almost certain" means to you.

$$z = \frac{\hat{p} - \mu_{\hat{p}}}{\sqrt{\frac{pq}{n}}}$$

$$-3 = \frac{0.75 - 0.80}{\sqrt{\frac{(0.80)(0.20)}{n}}}$$

$$n = \frac{(-3)^2(0.80)(0.20)}{(0.75 - 0.80)^2} = 576$$

26. Cigarettes 2006.

a) H_0: 20% of high school students smoke. ($p = 0.20$)
H_A: More than 20% of high school students smoke. ($p > 0.20$)

b) **Randomization condition:** The CDC randomly sampled 1,815 high school students.
10% condition: The sample of 1,815 students is less than 10% of all high school students.
Success/Failure condition: $np = (1{,}815)(0.20) = 363$ and $nq = (1815)(0.80) = 1452$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p = 0.20$ and $\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.20)(0.80)}{1815}} \approx 0.0094$.

c) We can perform a one-proportion z-test. The observed proportion of high school students who smoke is $\hat{p} = 0.23$ This proportion is about 3.17 standard deviations above the hypothesized proportion of smokers.

The P-value of this test is 0.0008.

d) If the proportion of students who smoke is actually 20%, the probability that a sample of this size would have a sample proportion of 23% or higher is 0.0008.

e) Since the P-value = 0.0008 is low, we reject the null hypothesis. There is strong evidence that greater than 20% of high school students smoked in 2006. The goal is not on track.

f) If the conclusion is incorrect, a Type I error has been made.

27. Crohn's disease.

a) **Independence assumption:** It is reasonable to think that the patients would respond to infliximab independently of each other.
Randomization condition: Assume that the 573 patients are representative of all Crohn's disease sufferers.
10% condition: 573 patients are less than 10% of all sufferers of Crohn's disease.
Success/Failure condition: $n\hat{p} = 335$ and $n\hat{q} = 238$ are both greater than 10.

Since the conditions are met, we can use a one-proportion z-interval to estimate the percentage of Crohn's disease sufferers who respond positively to infliximab.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\frac{335}{573}\right) \pm 1.960\sqrt{\frac{\left(\frac{335}{573}\right)\left(\frac{238}{573}\right)}{573}} = (54.4\%, 62.5\%)$$

b) We are 95% confident that between 54.4% and 62.5% of Crohn's disease sufferers would respond positively to infliximab.

c) 95% of random samples of size 573 will produce intervals that contain the true proportion of Crohn's disease sufferers who respond positively to infliximab.

28. Teen smoking 2006.

Randomization condition: Assume that the freshman class is representative of all teenagers. This may not be a reasonable assumption. There are many interlocking relationships between smoking, socioeconomic status, and college attendance. This class may not be representative of all teens with regards to smoking simply because they are in college. Be cautious with your conclusions!
10% condition: The freshman class is less than 10% of all teenagers.
Success/Failure condition: $np = (522)(0.23) = 120$ and $nq = (522)(0.77) = 402$ are both greater than 10.

Therefore, the sampling distribution model for the proportion of 522 students who smoke is Normal, with $\mu_{\hat{p}} = p = 0.23$, and standard deviation $\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.23)(0.77)}{522}} \approx 0.0184$.

30% is about 3.8 standard deviations above the expected proportion of smokers. According to the Normal model, the probability that more than 30% of these students smoke is very small. It is very unlikely that more than 30% the freshman class smokes.

29. Alcohol abuse.

$$ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.04 = 1.645\sqrt{\frac{(0.5)(0.5)}{n}}$$

$$n = \frac{(1.645)^2(0.5)(0.5)}{(0.04)^2}$$

$$n \approx 423$$

The university will have to sample at least 423 students in order to estimate the proportion of students who have been drunk with in the past week to within ± 4%, with 90% confidence.

30. Errors.

a) Since a treatment (the additive) is imposed, this is an experiment.

b) The company is only interested in a decrease in the percentage of cars needing repairs, so they will perform a one-sided test.

c) The independent laboratory will make a Type I error if they decide that the additive reduces the number of repairs, when it actually makes no difference in the number of repairs.

d) The independent laboratory will make a Type II error if they decide that the additive makes no difference in the number of repairs, when it actually reduces the number of repairs.

e) The additive manufacturer would consider a Type II error more serious. The lab claims that the manufacturer's product doesn't work, and it actually does.

f) Since this was a controlled experiment, the company can conclude that the additive is the reason that the cabs are running better. They should be cautious recommending it for all cars. There is evidence that the additive works well for cabs, which get heavy use. It might not be effective in cars with a different pattern of use than cabs.

31. Preemies.

a) **Randomization condition:** Assume that these kids are representative of all kids.
10% condition: 242 and 233 are less than 10% of all kids.
Independent samples condition: The groups are independent.
Success/Failure condition: $n\hat{p}$(preemies) = (242)(0.74) = 179, $n\hat{q}$(preemies) = (242)(0.26) = 63, $n\hat{p}$(normal weight) = (233)(0.83) = 193, and $n\hat{q}$(normal weight) = 40 are all greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will find a two-proportion z-interval.

$$\left(\hat{p}_N - \hat{p}_P\right) \pm z^* \sqrt{\frac{\hat{p}_N\hat{q}_N}{n_N} + \frac{\hat{p}_P\hat{q}_P}{n_P}} = (0.83 - 0.74) \pm 1.960\sqrt{\frac{(0.83)(0.17)}{233} + \frac{(0.74)(0.26)}{242}} = (0.017, 0.163)$$

We are 95% confident that between 1.7% and 16.3% more normal birth-weight children graduated from high school than children who were born premature.

b) Since the interval for the difference in percentage of high school graduates is above 0, there is evidence normal birth-weight children graduate from high school at a greater rate than premature children.

c) If preemies do not have a lower high school graduation rate than normal birth-weight children, then we made a Type I error. We rejected the null hypothesis of "no difference" when we shouldn't have.

32. Safety.

a) $\hat{p} \pm z^* \sqrt{\dfrac{\hat{p}\hat{q}}{n}} = (0.14) \pm 1.960 \sqrt{\dfrac{(0.14)(0.86)}{814}} = (11.6\%, 16.4\%)$

We are 95% confident that between 11.6% and 16.4% of Texas children wear helmets when biking, roller skating, or skateboarding.

b) These data might not be a random sample.

c)

$$ME = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.04 = 2.326 \sqrt{\frac{(0.14)(0.86)}{n}}$$

$$n = \frac{(2.326)^2(0.14)(0.86)}{(0.04)^2}$$

$$n \approx 408$$

If we use the 14% estimate obtained from the first study, the researchers will need to observe at least 408 kids in order to estimate the proportion of kids who wear helmets to within 4%, with 98% confidence.

(If you use a more cautious approach, estimating that 50% of kids wear helmets, you need a whopping 846 observations. Are you beginning to see why pilot studies are conducted?)

33. Fried PCs.

a) H_0: The computer is undamaged.
H_A: The computer is damaged.

b) The biggest advantage is that all of the damaged computers will be detected, since, historically, damaged computers never pass all the tests. The disadvantage is that only 80% of undamaged computers pass all the tests. The engineers will be classifying 20% of the undamaged computers as damaged.

c) In this example, a Type I error is rejecting an undamaged computer. To allow this to happen only 5% of the time, the engineers would reject any computer that failed 3 or more tests, since 95% of the undamaged computers fail two or fewer tests.

d) The power of the test in part c is 20%, since only 20% of the damaged machines fail 3 or more tests.

e) By declaring computers "damaged" if the fail 2 or more tests, the engineers will be rejecting only 7% of undamaged computers. From 5% to 7% is an increase of 2% in α. Since 90% of the damaged computers fail 2 or more tests, the power of the test is now 90%, a substantial increase.

34. Power.

a) Power will increase, since the variability in the sampling distribution will decrease. We are more certain of all our decisions when there is less variability.

b) Power will decrease, since we are rejecting the null hypothesis less often.

35. Approval 2007.

H_0 : George W. Bush's May 2007 disapproval rating was 66%. ($p = 0.66$)
H_A : George W. Bush's May 2007 disapproval rating was lower than 66%. ($p < 0.66$)

Independence assumption: One adult's response will not affect another's.
Randomization condition: The adults were chosen randomly.
10% condition: 1000 adults are less than 10% of all adults.
Success/Failure condition: $np = (1000)(0.66) = 660$ and $nq = (1000)(0.44) = 440$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p = 0.66$ and $\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.66)(0.44)}{1000}} \approx 0.017$.

We can perform a one-proportion z-test. The observed approval rating is $\hat{p} = 0.63$.

The value of z is – 2.00. Since the P-value = 0.023 is low, we reject the null hypothesis. There is strong evidence that President George W. Bush's May 2007 disapproval rating was lower than the 66% disapproval rating of President Richard Nixon.

36. Grade inflation.

H_0 : In 2000, 20%of students at the major university had a GPA of at least 3.5. ($p = 0.20$)
H_A : In 2000, more than 20%of students had a GPA of at least 3.5. ($p > 0.20$)

Independence assumption: Student's GPAs are independent of each other.
Randomization condition: The GPAs were chosen randomly.
10% condition: 1100 GPAs are less than 10% of all GPAs.
Success/Failure condition: $np = (1100)(0.20) = 220$ and $nq = (1100)(0.80) = 880$ are both greater than 10, so the sample is large enough.

The conditions have been satisfied, so a Normal model can be used to model the sampling distribution of the proportion, with $\mu_{\hat{p}} = p = 0.20$ and $\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.20)(0.80)}{1100}} \approx 0.0121$.

We can perform a one-proportion z-test. The observed approval rating is $\hat{p} = 0.25$.

Since the P-value is less than 0.0001, which is low, we reject the null hypothesis. There is strong evidence the percentage of students whose GPAs are at least 3.5 is higher in 2000 than in 1996.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.20 - 0.25}{\sqrt{\frac{(0.20)(0.80)}{1100}}}$$

$$z \approx 4.15$$

37. Name Recognition.

a) The company wants evidence that the athlete's name is recognized more often than 25%.

b) Type I error means that fewer than 25% of people will recognize the athlete's name, yet the company offers the athlete an endorsement contract anyway. In this case, the company is employing an athlete that doesn't fulfill their advertising needs.

Type II error means that more than 25% of people will recognize the athlete's name, but the company doesn't offer the contract to the athlete. In this case, the company is letting go of an athlete that meets their advertising needs.

c) If the company uses a 10% level of significance, the company will hire more athletes that don't have high enough name recognition for their needs. The risk of committing a Type I error is higher.

At the same level of significance, the company is less likely to lose out on athletes with high name recognition. They will commit fewer Type II errors.

38. Name Recognition, part II.

a) The 2% difference between the 27% name recognition in the sample, and the desired 25% name recognition may have been due to sampling error. It's possible that the actual percentage of all people who recognize the name is lower than 25%, even though the percentage in the sample of 500 people was 27%. The company just wasn't willing to take that chance. They'll give the endorsement contract to an athlete that they are convinced has better name recognition.

b) The company committed a Type II error. The null hypothesis (that only 25% of the population would recognize the athlete's name) was false, and they didn't notice.

c) The power of the test would have been higher if the athlete were more famous. It would have been difficult not to notice that an athlete had, for example, 60% name recognition if they were only looking for 25% name recognition.

39. NIMBY.

Randomization condition: Not only was the sample random, but Gallup randomly divided the respondents into groups.
10% condition: 502 and 501 are less than 10% of all adults.
Independent samples condition: The groups are independent.
Success/Failure condition: The number of respondents in favor and opposed in both groups are all greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will find a two-proportion z-interval.

$$\left(\hat{p}_1 - \hat{p}_2\right) \pm z^* \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} = \left(0.53 - 0.40\right) \pm 1.960\sqrt{\frac{(0.53)(0.47)}{502} + \frac{(0.40)(0.60)}{501}} = \left(0.07, 0.19\right)$$

We are 95% confident that the proportion of U.S. adults who favor nuclear energy is between 7 and 19 percentage points higher than the proportion that would accept a nuclear plant near their area.

40. Dropouts.

Randomization condition: Assume that these subjects are representative of all anorexia nervosa patients.
10% condition: 198 is less than 10% of all patients.
Success/Failure condition: The number of dropouts, 105, and the number of subjects that remained, 93, are both greater than 10, so the samples are both large enough.

Since the conditions have been satisfied, we will find a one-proportion z-interval.

$$\hat{p} \pm z^{*}\sqrt{\frac{\hat{p}\hat{q}}{n}} = \left(\tfrac{105}{198}\right) \pm 1.960\sqrt{\frac{\left(\frac{105}{198}\right)\left(\frac{93}{198}\right)}{198}} = (46\%, 60\%)$$

We are 95% confident that between 46% and 60% of anorexia nervosa patients will drop out of treatment programs. However, this wasn't a random sample of all patients. They were assigned to treatment programs rather than choosing their own. They may have had different experiences if they were not part of an experiment.