Review for Test 3

1. **From a random sample of 36 days in a recent year, the closing stock prices of Hasbro had a mean of $19.31. From past studies we know that the population standard deviation is $2.37.**

a. Should you use the *z*-distribution, or the *t*-distribution to construct a CI for the population mean for closing stock prices of Hasbro?

   Normal distribution (z) because the population standard deviation, σ, is given.

b. Construct the 90% and 95% confidence intervals for the population mean. Interpret the CIs.

   90% CI: Conditions: SRS, σ is known, and n >30.

   (#7 in your calculator)

   $$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} = 19.31 \pm 1.654 \frac{2.37}{\sqrt{36}} = (18.66, 19.96)$$

   We are 90% confident that the mean closing stock prices of Hasbro is between $18.66 and $19.96.

   95% CI: Conditions: SRS, σ is known, and n >30.

   $$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} = 19.31 \pm 1.96 \frac{2.37}{\sqrt{36}} = (18.54, 20.08)$$

   We are 95% confident that the mean closing stock prices of Hasbro is between $18.54 and $20.08.

c. Which interval is wider? What are the margins of errors?

   The 95% CI is wider. And it is suppose to be wider, because higher confidence level gives wider CI intervals.

   Margin of error for 90% confidence:
   $$1.654 \frac{2.37}{\sqrt{36}} = 0.65$$

   Margin of error for 95% confidence:
   $$1.96 \frac{2.37}{\sqrt{36}} = 0.774$$

2. **You randomly select 20 mortgage institutions and determine the current mortgage interest rate at each. The sample mean rate is 6.93% with a sample standard deviation of 0.42%. Find the 99% confidence interval for the population <u>mean</u> mortgage interest rate. Assume the interest rates are approximately normally distributed.**

   Since the population standard deviation, σ, is unknown, and n < 30, we need to use the t-interval: Conditions: SRS, population is normally distributed. Both satistfied.

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}} = 6.93 \pm 2.861 \frac{0.42}{\sqrt{20}} = (6.66, 7.20)$$

(#8 in your calculator)
We are 99% confident that the mean mortgage interest rate for ALL mortgage institutions is between 6.66% and 7.20%.

3. A biologist reports a CI of (2.1cm, 3.5cm) when estimating the <u>mean</u> height of a sample of seedlings. Find the sample mean and the margin of error.

The point estimator, the sample mean is exactly in the middle of the CI. The margin of error, $m,$ is the half of the interval.

```
              E                       E
      (-----------------------|-----------------------)
       2.1 cm                 x̄              3.5 cm
```
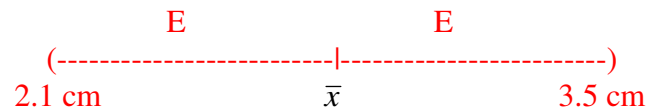
Thus, the sample mean is: $\bar{x} = \dfrac{2.1 + 3.5}{2} = 2.8 \ cm$

And the margin of error is: $E = 2.8 \ cm - 2.1 \ cm = 0.7 \ cm$

4. Determine the minimum required sample size if you want to be 95% confident that the sample mean is <u>within one unit</u> of the population mean when the standard deviation is 4.8. Assume that the population is normally distributed.

Margin of error: E = 1
$\sigma = 4.8$

$$n = \left(\frac{z^* \sigma}{E} \right)^2 = \left(\frac{2 \cdot 4.8}{1} \right)^2 = 92.16$$

Thus, the minimum required sample size is 93. (Always round UP!!)

5. A random sample of airfare prices in dollars for a one-way ticket from New York to Houston is given:

288, 290, 292, 295, 298, 300, 305, 305, 306, 307, 314, 320, 322, 326, 327

Assuming that the prices of tickets are normally distributed, find and interpret the 90% CI for the mean price of all tickets from New York to Houston. (The sample mean is $306.3, and the sample standard deviation is $12.97, t* = 1.761)

Since σ is not given, and n < 30, we need to use the t-test.
Conditions: SRS, normally distributed population: both are satisfied.

If you have a TI-83, 84, you can enter the list and then use STAT → TESTS → 8: TInterval and highlight Data.

90% CI: $\bar{x} \pm t^* \dfrac{s}{\sqrt{n}} = 306.3 \pm 1.761 \dfrac{12.97}{\sqrt{15}} = (300.43, 312.23)$

We are 90% confident that the mean price of ALL airfare prices for a one-way ticket from New York to Houston is between $300.43 and $312.23.

6. The following table shown is from a survey of randomly selected 900 U.S. adults.

Who are the more dangerous drivers?	
Teenagers	63%
People over 75	33%
No opinion	4%

a. Construct a 95% CI for the proportion of adults who think that teenagers are the more dangerous drivers. What's the margin of error?

Conditions: SRS, $n\hat{p} = 900(0.63) = 567 > 10$, $n\hat{p} = 900(0.63) = 567 > 10$. All satisfied.

$\hat{p} \pm z^* \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}} = 0.63 \pm 1.96 \sqrt{\dfrac{0.63(1-0.63)}{900}} = (0.598, 0.662)$

(A in your calculator)
We are 95% confident that the proportion of ALL adults who think that teenagers are the more dangerous drivers is between 59.8% and 66.2%.

The margin of error: $1.96 \sqrt{\dfrac{0.63(1-0.63)}{900}} = 0.032$

b. Construct a 95% CI for the proportion of adults who think that people over 75 are the more dangerous drivers. What is the margin of error?

Conditions: SRS, $n\hat{p} = 900(0.33) = 297 > 10$, $n\hat{p} = 900(1-0.33) = 603 > 10$. All satisfied.

$\hat{p} \pm z^* \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}} = 0.33 \pm 1.96 \sqrt{\dfrac{0.33(1-0.33)}{900}} = (0.3, 0.36)$

(A in your calculator)

We are 95% confident that the proportion of ALL adults who think that teenagers are the more dangerous drivers is between 59.8% and 66.2%.

The margin of error: $1.96\sqrt{\dfrac{0.33(1-0.33)}{900}} = 0.307$

7. You are running a political campaign and wish to estimate, with 99% confidence, the proportion of registered voters who will vote for your candidate. Your estimate must be accurate within 3% of the true population. Find the minimum sample size needed if

a. no preliminary estimation is available

When no preliminary estimation is available, use $\hat{p} = 0.5$

$$n = \left(\frac{z^*}{E}\right)^2 \hat{p}(1-\hat{p}) = \left(\frac{2.576}{0.03}\right)^2 (0.5)(1-0.5) = 1843.271$$

Thus, the minimum sample size required is 1844.

b. a preliminary estimation gives $\hat{p} = 0.31$.

$$n = \left(\frac{z^*}{E}\right)^2 \hat{p}(1-\hat{p}) = \left(\frac{2.576}{0.03}\right)^2 (0.31)(1-0.31) = 1577.102$$

Thus, the minimum sample size required is 1578.

8. A used car dealer says that the mean price of a 2002 Ford F-150 Super Cab is $18,800. You suspect this claim is too high. You find that a random sample of 14 similar vehicles has a mean price of $18,250 and a standard error of $1,250. Is there enough evidence to reject the dealer's claim at α=0.05? Assume the population is normally distributed. Is your conclusion the same at the 10% level? Can you reach the same conclusion by using a CI?

$H_0: \mu = 18800$

$H_a: \mu < 18800$

Conditions: SRS, normal distribution. Both satisfied.

Since σ is not given, and n < 30, we need to use a t-test. (#2 in your calculator menu)

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{1820 - 18800}{\frac{1250}{\sqrt{14}}} = -1.646$$

The p-value is 0.0618.

Since the p-value is greater than 0.05, the significance level, we cannot reject the null hypothesis. At the 5% level, we don't have enough evidence to reject the claim that the mean price of 2002 Ford F-150 Super Cab is $18,800.

At the 10% level, since the p-value is less than 10%, we can reject the null hypothesis. We have enough evidence to reject the claim that the mean price of 2002 Ford F-150 Super Cab is $18,800.

No, we can't use a CI to test the claim because it's not a two-tailed test. (We don't have an $\neq$ in the alternative hypothesis).

9. A company that makes cola drinks states that the mean caffeine content per one 12-ounce bottle of cola is 40 milligrams. You work as a quality control manager and are asked to test this claim. During your tests, you find that a random sample of 30 12-ounce bottles of cola has a mean caffeine content of 39.2 milligrams. From previous studied you know that the standard deviation of the population is 7.5 milligrams. Assume that the caffeine content is normally distributed. At α = 1% level, can you reject the company's claim? Can you reach the same conclusion by using a CI?

$H_0 : \mu = 40$

$H_a : \mu \neq 40$

Conditions: SRS, σ is known

Since σ is given, we can use a z-test. (#1 in your calculator)

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{39.2 - 40}{\frac{7.5}{\sqrt{30}}} = -0.584$$

p-value: 0.559

Conclusion: since the p-value is about 56%, we have no evidence to reject the null hypothesis. We don't have enough evidence to reject the claim that the mean caffeine level of all 12-ounce bottle of cola is 40 milligrams.

Yes, we can use a 99% confidence interval to test the claims since we have a two-sided test. (#7 in your calculator)

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} = 39.2 \pm 2.576 \frac{7.5}{\sqrt{30}} = (35.63, 42.727)$$

Since the 99% CI does contain the hypothesized value, 40, we cannot reject the null hypothesis at the 1% level.

10. A scientist estimates that the mean nitrogen dioxide level in West London is 28 parts per billion. You believe that the mean nitrogen dioxide level is higher. So, you determine the nitrogen dioxide levels for 36 randomly selected days. The results (in parts per billion) are listed below. At α = 10%, can you support the scientist's estimate? Can you reach the same conclusion by using a CI?

27, 29, 53, 31, 16, 47, 22, 17, 13, 46, 99, 15, 20, 17, 28, 10, 14, 9, 35, 29, 32, 67, 24, 31, 43, 29, 12, 39, 65, 94, 12, 27, 13, 16, 40, 62

(The sample mean is 32.86, and the sample standard deviation is 22.13)

Hypotheses: $H_0: \mu = 28$
$H_a: \mu > 28$

Since σ is not given, but n > 30, you can use the t-test.
If you have a Ti-83, 84, you can enter the list (STAT → EDIT → Edit, and then use STAT →
TESTS→ 2: T-test, and highlight Data.

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{32.86 - 28}{\frac{22.12}{\sqrt{36}}} = 1.318$$

p-value: 0.098
Conclusion: since p-value is less than α (0.098 < 10), strictly speaking the result is statistically significant. That is, we can reject the null hypothesis, although it's really a borderline case since 0.098 ≈ 0.10.
The decision is in your hands.

This would be my conclusion: We have little or no evidence against the null hypothesis. We have little or no evidence against the claim that the mean nitrogen dioxide level in West London is 28 parts per billion. Further investigation is necessary.

No, we can't use a CI interval here to test the claim because it's not a two-tailed test (we don't have ≠ in the alternative hypothesis.)

11. A medical researcher claims that 20% of adults in the U.S. are allergic to a medication. You think that less than 20% are allergic to a medication. In a random sample of 100 adults, 15% say they have such an allergy. Test the researcher's claim at α = 1% level. Can you reach the same conclusion using the appropriate confidence interval?

Hypotheses: $H_0: p = 0.20$
$H_a: p < 0.20$

Conditions: SRS, $np_0 = 100(0.20) = 20 > 10, \quad n(1 - p_0) = 100(1 - 0.20) = 80 > 10$, all satistfied.

Test statistic: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} = \frac{0.15 - 0.20}{\sqrt{\frac{0.20(1 - 0.20)}{100}}} = -1.25$

(You can use your calculator, STAT→TESTS→5:1-PropZTest. Use $x = n\hat{p} = 100(0.15) = 15$)
p-value: 0.1056

Conclusion: Since p-value > 10%, we have no evidence against the null hypothesis. We can't reject it. Based on our sample, we don't have enough evidence to reject the medical researcher's claim that 20% of adults in the U.S. are allergic to a medication.

No, we can't use a CI interval here to test the claim because it's not a two-tailed test (we don't have $\neq$ in the alternative hypothesis.

12.   USA TODAY reports that 5% of U.S. adults have seen an extraterrestrial being. You decide to test this claim and ask a random sample of 250 U.S. adults whether they have ever seen an extraterrestrial being. Of those surveyed, 8% reply yes. At $\alpha = 5\%$, is there enough evidence to reject the claim? Can you reach the same conclusion using the appropriate confidence interval?

Hypotheses:
$$H_0: p = 0.05$$
$$H_a: p \neq 0.05$$

Conditions: SRS, $np_0 = 250(0.05) = 12.5 > 10$, $n(1 - p_0) = 250(1 - 0.05) = 237.5 > 10$, all satisfied.

Test statistic: $z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1 - p_0)}{n}}} = \dfrac{0.08 - 0.05}{\sqrt{\dfrac{0.05(1 - 0.05)}{250}}} = 2.176$

(You can use your calculator, STAT→TESTS→5:1-PropZTest. Use $x = n\hat{p} = 250(0.08) = 20$)

p-value: 0.029

(If you want to find the p-value using the z-table, find the value for z = 2.176, and multiply it with 2 because it's a two-tailed test.)

Conclusion: Since p-value < 5%, we can reject the null hypothesis. We have good evidence against the USA TODAY's claim that 5% of U.S. adults have seen an extraterrestrial being.

Yes, we can use a 95% CI interval (the confidence level and the α must add up to 100%) because it's a two-tailed test (we have $\neq$ in our alternative hypothesis).

The 95% CI: $\hat{p} \pm z^* \sqrt{\dfrac{\hat{p}(1 - \hat{p})}{n}} = 0.08 \pm 1.96\sqrt{\dfrac{0.08(1 - 0.08)}{250}} = (0.046, 0.114)$

Since the claimed value, 0.05 is inside the 95% CI, at the 5% level we can't reject the null hypothesis claimed by USA TODAY that 5% of U.S. adults have seen an extraterrestrial being. Thus, we can reach the same conclusion what we got using the hypothesis test.

13.   **A medical researcher estimates that 55% of U.S. adults eat breakfast every day. In a random sample of 250 U.S. adults, 56.4% say that they eat breakfast every day.** At $\alpha = 5\%$, is there enough evidence to reject the researcher's claim?

Hypotheses:
$$H_0: p = 0.55$$
$$H_a: p \neq 0.55$$

Conditions: SRS, $np_0 = 250(0.55) = 137.5 > 10$, $n(1 - p_0) = 250(1 - 0.55) = 112.5 > 10$, all satisfied.

Test statistic: $z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1 - p_0)}{n}}} = \dfrac{0.564 - 0.55}{\sqrt{\dfrac{0.55(1 - 0.55)}{250}}} = 0.445$

(You can use your calculator, STAT→TESTS→5:1-PropZTest. Use $x = n\hat{p} = 250(0.55) = 141$)

p-value: 0.656

(If you want to find the p-value using the z-table, find the value for z = -0.445, and multiply it with 2 because it's a two-tailed test.)

Conclusion: Since p-value > 5%, we cannot reject the null hypothesis. We can't reject the claim that 55% of U.S. adults eat breakfast every day.

14. **Classify the two given samples as independent or paired samples:**

a. The SAT scores for 35 high school students who did not take an SAT preparation course.
The SAT scores for 40 high school students who did take an SAT preparation course.
independent

b. The SAT scores for 44 high school students
The SAT scores for the same 44 high school students after taking an SAT preparation course.
paired

c. The weights of 51 adults.
The weights of the same 51 adults after participating in a diet and exercise program for one month
paired

d. The weight of 40 females
The weight of 40 males
independent

e. The average speed of 23 powerboats using an old hull design
The average speed of 23 powerboats using a new hull design
independent

f. The fuel mileage of 10 cars
The fuel mileage of the same 10 cars using a fuel additive
paired

g. Heart rate of 10 people before exercising
Heart rate of the same10 people after exercising
Paired

15. **Consumer Reports tested the stopping distances (in feet) of 10 randomly selected Firestone Winterfire tires and 12 Michelin XM-S Alpin tires when traveling on ice at 15 mph. Results are shown below.** Based on the results, can you conclude that there is a difference in the mean stopping distances of the two types of tires? Use 10% significance level. State the two hypothesis, and write your conclusion in context based on the p-value and CI.

p-value: 0.163   90% CI: (-8.798, 0.798)

Hypotheses:
$$H_0 : \mu_{Firestone} = \mu_{Michelin}$$
$$H_a : \mu_{Firestone} \neq \mu_{Michelin}$$

Conclusion: Since p-value > 10%, we cannot reject the null hypothesis. We can't reject the claim that the mean stopping distances of the two types of tires are the same. So we can't conclude that there is a difference in the stopping distances of the two tires.
Based on the CI, we again, can't reject the null hypothesis because 0, the claimed value is inside the confidence interval. Same conclusion as above.

16. **A manufacturer claims that the calling range (in feet) of its 2.4 GHz cordless phone is greater than that of its leading competitor. You perform a study using 14 randomly selected phones from the manufacturer and 16 randomly selected phones from its competitor.** At the 5% significance level, can you support the manufacturer's claim? Assume the populations are normally distributed.
State the claims, and write your conclusion in context based on the given p-value.

Manufacturer: sample mean = 1275 ft, sample s.d. = 45 ft, sample size = 14
Competitors: sample mean = 1250 ft, sample s.d. = 30 ft, sample size = 16
p-value: 0.046

Hypotheses:
$$H_0 : \mu_{manufacturere} = \mu_{competitor}$$
$$H_a : \mu_{manufacturer} > \mu_{competitor}$$

Conclusion: Since p-value < 5%, we can reject the null hypothesis. We can reject the claim that the mean calling ranges (in feet) are the same for both phones.

17. **A physician claims that an experimental medication increases an individual's heart rate. Twelve test subjects are randomly selected, and the heart rate of each is measured. The subjects are then injected with the medication and, after 1 hour, the heart rate of each is measured again.** The results are show below.

Before	72	81	76	74	75	80	68	75	78	76	74	77
After	73	80	79	76	76	80	74	77	75	74	76	78

State the hypothesis, and your conclusion in context using 5% significance level.
p-value: 0.086

This is paired data, since it's a "before-after" experiment. Since we want to "prove" that the medication increases the heart rate, the difference of "Before – After" should be negative. ("Before" is smaller, and "After" is a greater number)

Hypotheses:
$$H_0: \mu_d = 0$$
$$H_a: \mu_d < 0$$

Conclusion: Since p-value > 5%, we cannot reject the null hypothesis. We cannot reject the claim that the mean of the differences between the heart rates is 0.

18. **The table shows the gas mileage (in miles per gallon) of 10 cars with and without using a fuel additive. At the 5% significance level, is there enough evidence to conclude that the fuel additive improved gas mileage?**

| Gas mileage without additive | 35.8 | 37.7 | 39.4 | 36.8 | 36.6 | 33.7 | 38.4 | 37.3 | 35.7 | 36.7 | 37.1 | 38.3 |
| Gas mileage with additive | 36.2 | 39.8 | 40.1 | 39.3 | 36.9 | 34.5 | 38.8 | 37.3 | 36.1 | 37.9 | 37.0 | 38.6 |

p-value = 0.0036

This is paired data, again. Since we want to "prove" that the fuel additive improves gas mileage, the difference of "Without additive – with additive" should be negative. ("With additive" is smaller, and "with additive" is a greater number)

Hypotheses:
$$H_0: \mu_d = 0$$
$$H_a: \mu_d < 0$$

Conclusion: Since p-value < 5%, we can reject the null hypothesis. We can reject the claim that the mean of the differences in mileage is 0.

19. **A medical research team conducted a study to test the effect of a cholesterol-reducing medication. At the end of the study, the researchers found that of the 4700 subjects who took the medication, 301 died of heart disease. Of the 4300 subjects who took the placebo, 357 died of heart disease.** At the 1% significance level, can you conclude that the death rate is lower for those who took the medication than for those who took the placebo?
State the hypotheses, find the test statistic (z), and the p-value. State your conclusion in context.

This is a test to compare two populations' proportions.

Hypotheses:
$$H_0: p_{medication} = p_{placebo}$$
$$H_a: p_{medication} < p_{placebo}$$

Test statistic: you can use #6 in your calculator, or the formulas:

$$z = \frac{(\hat{p}_{medication} - \hat{p}_{placebo}) - 0}{\sqrt{\tilde{p}(1-\tilde{p})\left(\dfrac{1}{n_{medication}} + \dfrac{1}{n_{placebo}}\right)}}$$

where

$$\tilde{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{301 + 357}{4700 + 4300} = 0.073$$

$$z = \frac{(\hat{p}_{medication} - \hat{p}_{placebo}) - 0}{\sqrt{\tilde{p}(1-\tilde{p})\left(\dfrac{1}{n_{medication}} + \dfrac{1}{n_{placebo}}\right)}} = \frac{\left(\dfrac{301}{4700} - \dfrac{357}{4300}\right) - 0}{\sqrt{0.073(1-0.073)\left(\dfrac{1}{4700} + \dfrac{1}{4300}\right)}} = -3.45$$

p-value = 2.75×10^{-4} = 0.000275

If you are using the z-table, look up -3.45.

Conclusion: since the p-value < 1%, we can reject the null hypothesis. That is, we can reject the claim that the death rates are the same for the medication and the placebo groups.

20. **In a random survey of 1500 adults in California and 1000 adults in Oregon, you find that the percentages of smokers are 16.4% and 22.4%, respectively.** At the 5% significance level, is there enough evidence to conclude that the proportion of adults who are smokers is lower in California than in Oregon?
State the hypotheses, find the test statistic (z), and the p-value. State your conclusion in context.

This is a test to compare two populations' proportions.

Hypotheses: $\begin{aligned} H_0 &: p_{CA} = p_{OR} \\ H_a &: p_{CA} < p_{OR} \end{aligned}$

Test statistic: you can use #6 in your calculator, or the formulas:
$$z = \frac{(\hat{p}_{CA} - \hat{p}_{OR}) - 0}{\sqrt{\tilde{p}(1-\tilde{p})\left(\dfrac{1}{n_{CA}} + \dfrac{1}{n_{OR}}\right)}}$$

where

$$\tilde{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{246 + 224}{1500 + 1000} = 0.188$$

$$x_1 = 1500(0.164) = 246$$

$$x_2 = 1000(0.224) = 224$$

$$z = \frac{(\hat{p}_{CA} - \hat{p}_{OR}) - 0}{\sqrt{\tilde{p}(1 - \tilde{p})\left(\dfrac{1}{n_{CA}} + \dfrac{1}{n_{OR}}\right)}} = \frac{(0.164 - 0.224) - 0}{\sqrt{0.188(1 - 0.188)\left(\dfrac{1}{1500} + \dfrac{1}{1000}\right)}} = -3.76$$

p-value = 8.44×10^{-5} = 0.0000844

If you are using the z-table, look up -3.76.

Conclusion: since the p-value < 1%, we can reject the null hypothesis. That is, we can reject the claim that the proportion of smokers in CA and OR are the same.