Ch. 7 + 8.1-8.5

7 m/c   - 10 pt.

2 f/R   - 15 pt

1. True or False? Explain.

    → n is in denom.

    T a. For a fixed confidence level, when the sample size increases, the length of the confidence interval for a population mean decreases.

    F b. The z score corresponding to a 98 percent confidence level is 1.96.

    .01 ⟨—.98→⟩ .01
    ↳ 2.326

    T c. The best point estimate for the population mean is the sample mean.

    F d. The larger the level of confidence, the shorter the confidence interval.

    e. The margin of error can be computed from $\pm z^* \cdot \dfrac{\sigma}{\sqrt{n}}$ ← T if one sample w/ known pop. st. dev.

    F f. A statement contradicting the claim in the null hypothesis is classified as the power. alternate hyp.

    F g. If we want to claim that a population parameter is different from a specified value, this situation can be considered as a one-tailed test. two-tailed ≠

    F h. In the P-value approach to hypothesis testing, if the P-value is less than a specified significance level, we fail to reject the null hypothesis.
    p-value < α

    T i. A 90% confidence interval for a population parameter means that if a large number of confidence intervals were constructed from repeated samples, then on average, 90% of these intervals would contain the true parameter.

    T j. The point estimate of a population parameter is always at the center of the confidence interval for the parameter.
    ↑ statistic

2. Suppose that prior to conducting the coin-flipping experiment, we suspect that the coin is fair. How many times would we have to flip the coin in order to obtain a 90% confidence interval of width of at most 0.1 for the probability of flipping a head?

$$\overset{.05}{\overbrace{}}$$

$$ME = .05$$

$$.1$$
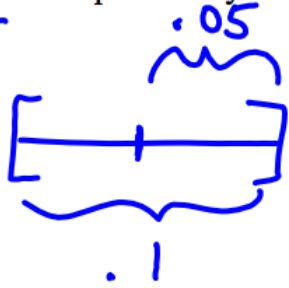
$$p \text{ not given} \rightarrow \text{use } p = .5$$

$$ME > z^* \sqrt{\frac{p(1-p)}{n}}$$

$$.05 > 1.645 \sqrt{\frac{.5(.5)}{n}}$$

$$90\% \ CI \Rightarrow z^* = 1.645$$

$$.05 > \frac{1.645 \sqrt{.5(.5)}}{\sqrt{n}}$$

$$\sqrt{n} > \frac{1.645(.5)}{.05}$$

$$n > 270.6 \qquad \boxed{n = 271}$$
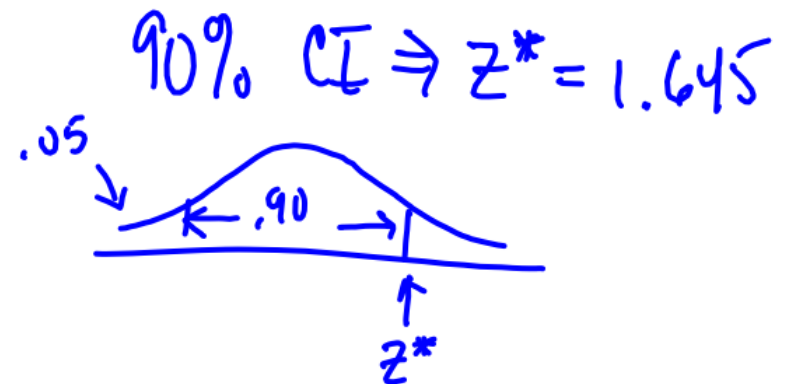
3. A certain beverage company is suspected of under filling its cans of soft drink. The company advertises that its cans contain, on the average, 12 ounces of soda with standard deviation 0.4 ounce. Compute the probability that a random sample of 50 cans produces a sample mean fill of 11.9 ounces or less.

$\mu = 12 \quad \sigma = .4$

$n = 50$

$$P(\bar{x} < 11.9) = \text{normal cdf} \left(-9999, 11.9, 12, \frac{.4}{\sqrt{50}} \right)$$

$H_0: \mu = 12$

$H_a: \mu < \cancel{12}$

$$Z = \frac{11.9 - 12}{.4/\sqrt{50}} = -1.7678$$

$$P(z < -1.7678) = \boxed{.0385}$$

$\text{normalcdf} (-9999, -1.7678)$

$\nearrow$

$0, 1$

4. A Brinell hardness test involves measuring the diameter of the indentation made when a hardened steel ball is pressed into material under a standard test load. Suppose that the Brinell hardness is determined for each specimen in a sample of size 50, resulting in a sample mean hardness of 64.3 and a sample standard deviation of 6.0. Calculate a 99% confidence interval for the true average Brinell hardness for material specimens of this type.

$$n = 50 \quad , \quad \bar{X} = 64.3 \quad S = 6.0$$

$$99\% \ CI$$

$$t^* = invT(.995, 49)$$
$$\text{gT}(.995, 49) \quad > \quad 2.68$$

$$64.3 \pm 2.68 \ \frac{6.0}{\sqrt{50}}$$

$$64.3 \pm 2.274$$

$$[62.026, 66.574]$$

5. The shear strength of anchor bolts has a standard deviation of 1.30. Assuming that the distribution is normal, how large a sample is needed to determine with a precision of ±0.5 the mean length of the produced needles to 99% confidence?

$n = ?$ $ME = .5$

$$\sigma = 1.3$$

$$z^* = \text{invNorm}(.995) = 2.576$$

$$.5 > 2.576 \frac{1.3}{\sqrt{n}}$$ $$ME > z^* \cdot \frac{\sigma \text{ or } s}{\sqrt{n}}$$

$$\sqrt{n} > \frac{2.576(1.3)}{.5}$$

$$n > 44.86$$ $$\boxed{n = 45}$$

6. The true average tread lives of two competing brands of radial tires (brand X and brand Y) are known to be normally distributed. The standard deviation of brand X tires is known to be 2200, and the standard deviation of brand Y tires is known to be 1900. A sample of 45 brand X tires results in a sample mean of 42,500 and sample standard deviation of 2450. A sample of 45 brand Y tires results in a sample mean of 40,400 and sample standard deviation of 2150. Find a 95% confidence interval for the difference in the true means, mean of X minus mean of Y.

z^* $\longrightarrow$ $\sigma_x = 2200$ $\sigma_y = 1900$

$n_x = 45$ $n_y = 45$

$\bar{x} = 42500$ $\bar{y} = 40400$

$z^*_{.95}$ $\boxed{1.96}$

$$(42500 - 40400) \pm \sqrt{\frac{2200^2}{45} + \frac{1900^2}{45}}$$

$$2100 \pm 849.327$$

$$[1250.67, 2949.33]$$

7. A sample of 97 Duracell batteries produces a mean lifetime of 10.40 hours and standard deviation 4.83 hours. A sample of 148 Energizer batteries produces a mean lifetime of 9.26 hours and a standard deviation of 4.68 hours. At a 5% significance level, can we assert that the average lifetime of Duracell batteries is greater than the average lifetime of Energizer batteries?

n_1 $\bar{x}_1$ s_1

$n_2 = 148$ $\bar{x}_2 = 9.26$ $s_2 = 4.68$ $\alpha = .05$

$H_0: \mu_1 = \mu_2$.

$H_a: \mu_1 > \mu_2$

1.66
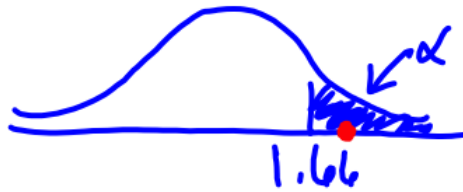
$df = 96$

$InvT(.95, 96) = 1.66$

two sample t-test.

$t = \dfrac{(10.40 - 9.26)}{\sqrt{\dfrac{4.83^2}{97} + \dfrac{4.68^2}{148}}} = \boxed{1.83}$

pvalue: $p(t > 1.83) = tcdf(1.83, 99999, 96)$

$1 - pt(1.83, 96)$

Reject H_0

Because the pvalue of .035 is less than 5%, we reject the null hyp. which states there is no difference in favor of saying Duracell has greater lifetime than Energizer.

$> .035 < \alpha$

8. In a sample of 539 households from a certain Midwestern city, it was found that 133 of these households owned at least one firearm. Give a 99% confidence interval for the percentage of families in this city who own firearms.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$z^* = \text{invNorm}(.995)$$
$$= 2.576$$

$$\frac{133}{539} \pm 2.576 \sqrt{\frac{\frac{133}{539}\left(1-\frac{133}{539}\right)}{539}}$$

$$\frac{133}{539} \pm .0478$$

$$.24675 \pm .0478$$

$$[.199, .295]$$

9. In an experiment to study the effects of illumination level on performance, subjects were timed for completion in both a low light level and high light level. The results are below.
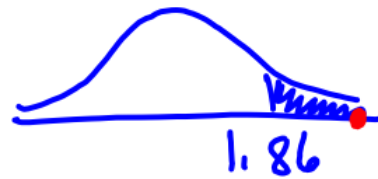
	Subject								
	1	2	3	4	5	6	7	8	9
Low Light	26	29	32	26	21	41	25	25	27
High Light	18	21	23	20	20	25	16	16	25

Low - High   8>0   8   9   6   1   16   9   9   2

Can you say with 95% certainty that the average completion time is lower in high light?

matched pairs t-test

$H_0: \mu_D = 0$

$H_a: \mu_D > 0$

1.86

$n = 9$

$\bar{x}_D = 7.556$

$df = 8$

$S_D = 4.39$

$t^* = (.95, 8) = 1.86$

$t = \dfrac{7.556 - 0}{4.39/\sqrt{9}} = 5.16$

pvalue $p(t > 5.16)$

$tcdf(5.16, 9999, 8)$

$1 - pt(5.16, 8) = .0004 < \alpha$

Based on $\alpha = .05$ (95% certainty)
I can reject the null hyp which states there is no difference
in favor of saying

10. A study of the ability of individuals to walk in a straight line reported the accompanying data on cadence (strides per second for a sample of $n = 20$ randomly selected men).

.95	.85	.92	.95	.93	.86	1.00	.92	.85	.81
.78	.93	.93	1.05	.93	1.06	1.06	.96	.81	.96

$\bar{X} = .9255$

Assuming the standard deviation of the population is 0.08: σ

   a. Find a 99% confidence interval for the mean cadence of the population.
   b. Test the hypothesis that the mean cadence for the population is less than 0.97 at the 5% significance level.

a. $.9255 \pm 2.576\left(\dfrac{.08}{\sqrt{20}}\right) = [.8795, .9715]$

b. $H_0 : \mu = .97$ .05
   $H_a : \mu < .97$ -1.645

$z = \dfrac{.9255 - .97}{.08/\sqrt{20}} = -2.488$

p value $p(z < -2.488) = .006 < \alpha \Rightarrow$ reject H_0

11. Bottles of a popular cola drink are supposed to contain 300 ml of cola. There is some variation from bottle to bottle because the filling machinery is not perfectly precise. The distribution of the contents is normal with standard deviation of 3 ml. A student who suspects that the bottler is under-filling measures the contents of six bottles. The results are:

$\mu = 300$

$\delta = 3$

| 299.4 | 297.7 | 301.0 | 298.9 | 300.2 | 297.0 |

$\bar{X} = 299.0333$ $n = 6$

Is this convincing evidence that the mean contents of cola bottles is less than the advertised 300 ml? Test at the 5% significance level.

One Sample mean z test

$H_0: \mu = 300$

$H_a: \mu < 300$

-1.645

$Z = \dfrac{299.0833 - 300}{3/\sqrt{6}} = -.789$
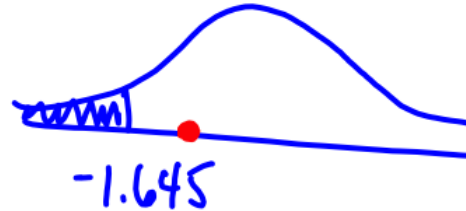
pvalue: $p(z < -.789) = .215 > \alpha$

Fail to reject H_0 which claims the drinks contain 300 ml of cola.

12. The guidance office of a school wants to test the claim of an SAT test preparation company that students who complete their course will improve their SAT Math score by at least 50 points. Ten members of the junior class who have had no SAT preparation but have taken the SAT once were selected at random and agreed to participate in the study. All took the course and re-took the SAT at the next opportunity. The results of the testing indicated:

Student	1	2	3	4	5	6	7	8	9	10
Before	475	512	492	465	523	560	610	477	501	420
After	500	540	512	530	533	603	691	512	489	458

$\bar{X}_D = -33.3$

bef - aft   -25 -28 -20 -65 -10 -43 -81   -35 12 -38

Is there sufficient evidence to support the prep course company's claim that scores will improve at the 1% level of significance?

$S_D = 26.39$

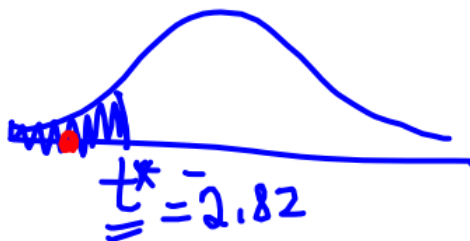higher afterscore

match pairs t-test

$H_0 : \mu_D = 0$

$H_a : \mu_D < 0$

$t^* = \text{InvT}(.01, 9) = -2.82$

$t^* = -2.82$

$t = \dfrac{-33.3 - 0}{26.39/\sqrt{10}} = -3.99$

pvalue: $P(t < -3.99) = .0016 < \alpha$ Reject H_0

13. A random sample of 200 freshmen and 100 seniors at Upper Wabash Tech are asked whether they agree with a plan to limit enrollment in crowded majors as a way of keeping the quality of instruction high. Of the students sampled, 160 freshmen and 20 seniors opposed the plan. We want to determine if there is any difference between the proportion of freshmen who oppose the plan and the proportion of seniors who oppose it.

   a. Formulate the null and alternative hypothesis.
   b. Compute the appropriate test statistic.
   c. Determine the p-value.
   d. Do you reject H_o or fail to reject H_o? Explain.
   e. Describe your results for someone who has no training in statistics.
   f. Find a 95% confidence interval for the difference between the population proportions.
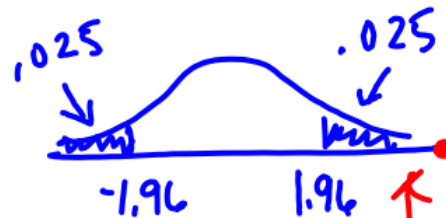
2 prop z-test

$$(.8 - .2) \pm 1.96 \sqrt{\frac{.8(.2)}{200} + \frac{.2(.8)}{100}}$$

Freshman: $n_1 = 200$ $\hat{p}_1 = \frac{160}{200} = .8$

Seniors: $n_2 = 100$ $\hat{p}_2 = \frac{20}{100} = .2$

$H_0 : P_1 = P_2$

$H_a : P_1 \neq P_2$

.025   .025

-1.96   1.96

$\alpha = .05$

$$z = \frac{.8 - .2}{\sqrt{\frac{.8(.2)}{200} + \frac{.2(.8)}{100}}} = 12.25$$

pvalue: $2p(z > 12.25) \approx 0$

14. It is fourth down and a yard to go for a first down in an important football game. The football coach must decide whether to go for the first down or punt the ball away. The null hypothesis is that the team will not get the first down if they go for it. The coach will make a Type I error by doing what?

A type I error occurs when one rejects the null hypothesis when it is true.

H_0: team will not get 1st down if they go for it.

they go for the 1st down + dont get 1st down

15. In a recent publication, it was reported that the average highway gas mileage of tested models of a new car was 33.5 mpg and approximately normally distributed. A consumer group conducts its own tests on a simple random sample of 12 cars of this model and finds that the mean gas mileage for their vehicles is 31.6 mpg with a standard deviation of 3.4 mpg.
    a. Perform a test to determine if these data support the contention that the true mean gas mileage of this model of car is different from the published value.
    b. Perform a test to determine if these data support the contention that the true mean gas mileage of this model of car is less than the published value.
    c. Explain why the answers to part a and part b are different.
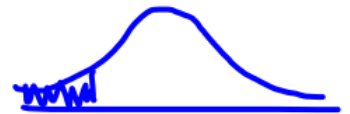
a. $H_0: \mu = 33.5$

$H_a: \mu \neq 33.5$

b. $H_0: \mu = 33.5$

$H_a: \mu < 33.5$

$n = 12$

$\bar{X} = 31.6$

$S = 3.4$

$t = \dfrac{31.6 - 33.5}{3.4/\sqrt{12}} = -1.936$

a. pvalue: $2p(t < -1.936) = .0790 > .05$ Fail to rej.

b. pvalue: $p(t < -1.936) = .0395 < .05$ Reject H_0

16. A random sample of size 36 selected from a normal distribution with $\sigma = 4$ has $\bar{x} = 75$. A second random sample of size 25 selected from a different normal distribution with $\sigma = 6$ has $\bar{x} = 85$. Is there a significant difference between the two population means at the 5% level of significance?

2 sample mean z-test $df = 24$

$H_0 : \mu_1 = \mu_2$

$\cdot \; n_1 = 36 \quad \sigma_1 = 4 \quad \bar{X}_1 = 75$

$H_a : \mu_1 \neq \mu_2$

$n_2 = 25 \quad \sigma_2 = 6 \quad \bar{X}_2 = 85$

$.025$

$-1.96 \qquad 1.96$

$z = \dfrac{75 - 85}{\sqrt{\dfrac{4^2}{36} + \dfrac{6^2}{25}}} = -7.28$

pvalue: $2 \cdot p(z < -7.28) = 0$

Reject Ho in favor of saying there is a difference in the populations.

17. A study was conducted to determine whether remediation in basic mathematics enabled students to be more successful in an elementary statistics course. (Success here means C or better.) Here are the results of the study:

①   ②

	Remedial	Non-remedial
Sample size	100	40
# of successes	70	16

$\hat{P}_1 = \frac{70}{100}$ $n_1 = 100$
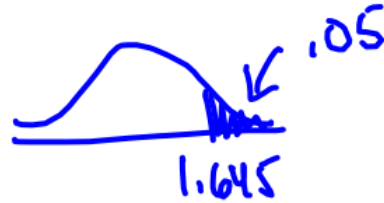
$\hat{P}_2 = \frac{16}{40}$ $n_2 = 40$

Test, at the 5% level, whether the remediation helped the students to be more successful.

2 sample prop z-test.

$H_0: P_1 = P_2$

$H_a: P_1 > P_2$

.05

1.645

$z = \dfrac{.7 - .4}{\sqrt{\dfrac{.7(.3)}{100} + \dfrac{.4(.6)}{40}}} = 3.333$

$p(z > 3.33) = .00043$

Reject H_0

18. A preacher would like to establish that of people who pray, less than 80% pray for world peace. In a random sample of 110 persons who pray, 77 of them said that when they pray, they pray for world peace. Test at the 10% level.
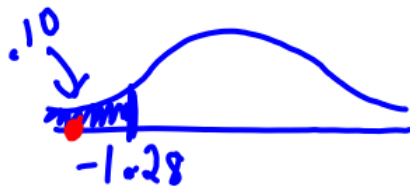
$\alpha = .10$

One sample prop z test

$H_0: p = .8$

$H_a: p < .8$

$\hat{p} = \frac{77}{110} = .7$

InvNorm(.10)

-1.28

$z = \frac{.7 - .8}{\sqrt{\frac{.8(.2)}{110}}} = -2.62$

pvalue: $p(z < -2.62)$

normalcdf$(-99999, -2.62, 0, 1)$

pnorm(-2.62)

$= .0044 < \alpha$

Reject H_0 in favor of saying less than 80% pray for world peace

PT 3

19) Mars Inc. claims that they produce M&Ms with the following distributions:

Brown	20%	Red	25%	Yellow	25%
Orange	5%	Green	15%	Blue	10%

Exp

22.2	27.75	27.75
5.55	16.65	11.1

mult by

A bag of M&Ms was randomly selected from the grocery store shelf, and the color counts were:

Brown	25	Red	23	Yellow	21
Orange	13	Green	15	Blue	14

← OBS.

$n = 111$

Using the χ^2 goodness of fit test ($\alpha = 0.05$) to determine if the proportion of M&Ms is what is claimed. Select the [test statistic, p-value, Decision to Reject (RH$_0$) or Failure to Reject (FRH$_0$)].

$df = 5$

$$\sum \frac{(0-E)^2}{E} = \chi^2 \qquad \text{pvalue: } p\,(\chi^2 > \#\,)$$

$$\chi^2 = \frac{(25-22.2)^2}{22.2} + \frac{(23-27.75)^2}{27.75} + \frac{(21-27.75)^2}{27.75} + \frac{(13-5.55)^2}{5.55}$$

$$+ \frac{(15-16.65)^2}{16.65} + \frac{(14-11.1)^2}{11.1} = 13.729$$

$$\text{pvalue: } p\,(\chi^2 > 13.729) = \chi^2 cdf\,(13.729, 99999, 5) = .0174$$

$$1 - 10 = A$$

Hypothesis tests:

Test	Null Hypothesis	Test Statistic
One-sample z-test for means	$\mu = \mu_o$	$z = \dfrac{\bar{x} - \mu_o}{\dfrac{\sigma}{\sqrt{n}}}$
One-sample t-test for means	$\mu = \mu_o$	$t = \dfrac{\bar{x} - \mu_o}{\dfrac{s}{\sqrt{n}}}$; df = n–1
Matched Pairs t-test	$\mu_D = \mu_{D_0}$	$t = \dfrac{\bar{x}_D - \mu_D}{s / \sqrt{n}}$; df = n – 1
One-sample z-test for proportions	$p = p_o$	$z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}}$
Two-sample t-test for means	$\mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$	$t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$; df=min(n1,n2)-1
Two-sample z-test for proportion	$p_1 - p_2 = 0$ or $p_1 = p_2$	$z = \dfrac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\left(\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}\right)}}$
χ^2 Goodness of fit test	no change	$\chi^2 = \sum \dfrac{(\text{observed} - \text{expected})^2}{\text{expected}}$

Confidence Intervals

One-sample z-test: $\quad \bar{x} \pm z * \dfrac{\sigma}{\sqrt{n}}$

Two-proportion z-test: $(\hat{p}_1 - \hat{p}_2) \pm z * \sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

One-sample t-test: $\quad \bar{x} \pm t * \dfrac{s}{\sqrt{n}}$

One-proportion z-test: $\hat{p} \pm z * \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

Two-sample z-test: $\quad (\bar{x}_1 - \bar{x}_2) \pm z * \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$

Two-sample t-test: $\quad (\bar{x}_1 - \bar{x}_2) \pm t * \sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$