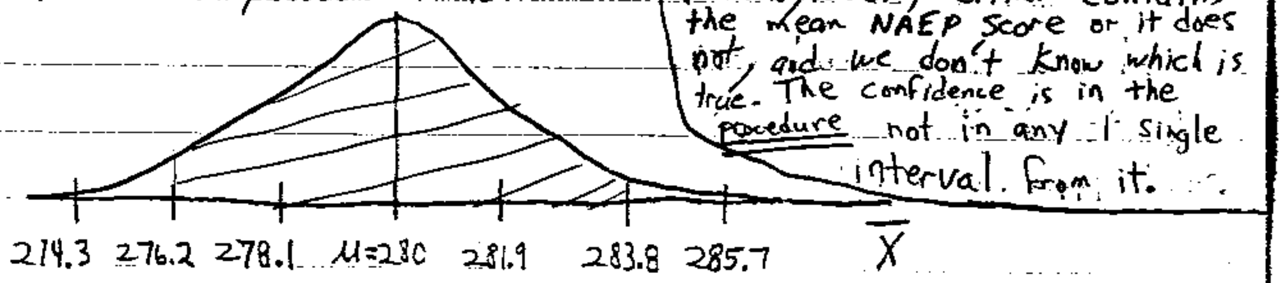
Lab 2 Soln

(6.2) No, we are 95% confident that (267.8, 276.2) contains the mean NAEP score.
95% confident means that if we repeat the procedure {① take another SRS ② construct a 95% CI from new sample} many times 95% of the intervals would contain the mean NAEP score.

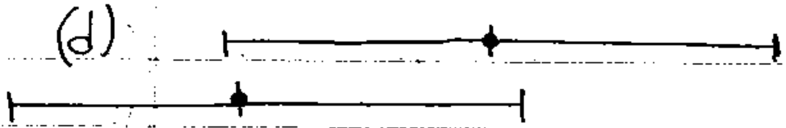
(267.8, 276.2) either contains the mean NAEP score or it does not, and we don't know which is true. The confidence is in the procedure not in any 1 single interval from it.

(6.3) (a) $\sigma/\sqrt{n} = 60/\sqrt{1000} \approx 1.9$

(b)



(d)



(c) $m = 2 * 1.9 = 3.8$

(e) $\approx 95\%$

(6.5) (a) Use a "back-to-back" Stem and leaf plot to order data quickly.

unordered leaves

42	7	24
	7	
	8	
69	8	69
31	9	13
68	9	68
333240	10	023334
785	10	578
22211444	11	11222444
98	11	89
0	12	0
8	12	8
20	13	02

"outlying observations"

ordered leaves

The data display an approximate normal shape. There are 2 small observations that are outlying.

(b) From JMP, $\bar{X} = 105.8$.

$\bar{X} \pm z^* \frac{\sigma}{\sqrt{n}}$

$105.8 \pm 2.576 \frac{15}{\sqrt{31}}$

105.8 ± 6.9

$(98.9, 112.7)$

(C) You want to make inference on 7th grade girls in the school district. You have a sample of 7th grade girls from only one of the schools in the district. This school could be the "best" one in the district. Thus, your results would be biased.

Lab 2 Soln.

Additional Part for 6.5: $\bar{x} \pm z^* \sigma \sqrt{1+\frac{1}{n}}$

$105.8 \pm 2.576 * 15 \sqrt{1+\frac{1}{31}}$

$105.8 \pm 38.64 * 1.016$

105.8 ± 39.3

$(66.5, 145.1)$

6.8 (a) $\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$

$275 \pm 1.96 \frac{60}{\sqrt{1077}}$

275 ± 3.6

$(271.4, 278.6)$

(b) 90% CI

$275 \pm 1.645 \frac{60}{\sqrt{1077}}$

275 ± 3.0

$(272, 278)$

99 % CI

$275 \pm 2.576 \frac{60}{\sqrt{1077}}$

275 ± 4.7

$(270.3, 279.7)$

(c)

Confidence level	Margin of Error
90%	3.0
95%	3.6
99%	4.7

Increasing confidence level, increases margin of error.

6.9 (a) Same as 6.8(a) -- $(271.4, 278.6)$.

(b) $275 \pm 1.96 \frac{60}{\sqrt{250}}$

275 ± 7.4

$(267.6, 282.4)$

(c) $275 \pm 1.96 \frac{60}{\sqrt{4000}}$

275 ± 1.9

$(273.1, 276.9)$

(d)

n	Margin of Error
250	7.4
1077	3.6
4000	1.9

Increasing n decreases margin of error.

Lab 2 Soln

(6.12) $n = \left(\frac{z^* \sigma}{m}\right)^2 = \left(\frac{2.576 * 15}{5}\right)^2 = \boxed{60}$

(6.18) (a) The authors wish to draw inference on all US citizens. However, their sample can only be fully trusted for inference on all citizens in Indianapolis (listed in the phone book).

(b)

$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$:	Food Stores	Mass merchandisers	Pharmacies
	$18.67 \pm 1.96 \frac{24.95}{\sqrt{201}}$	$32.38 \pm 1.96 \frac{33.37}{\sqrt{201}}$	$48.60 \pm 1.96 \frac{35.62}{\sqrt{201}}$
	18.67 ± 3.45	32.38 ± 4.61	48.60 ± 4.92
	(15.22, 22.12)	(27.77, 36.99)	(43.68, 53.52)

(c) Yes, for pharmacies the 95% CI for μ is entirely above that for both food stores and mass merchandisers. (i.e., $22.12 < 43.68$ AND $36.99 < 43.68$).

Additional Part For 6.18: $\bar{x} \pm z^* \sigma \sqrt{1 + \frac{1}{n}}$

Food Stores: $18.67 \pm 1.96 * 24.95 \sqrt{1 + \frac{1}{201}} = 18.67 \pm 49.02 = (0, 67.69)$

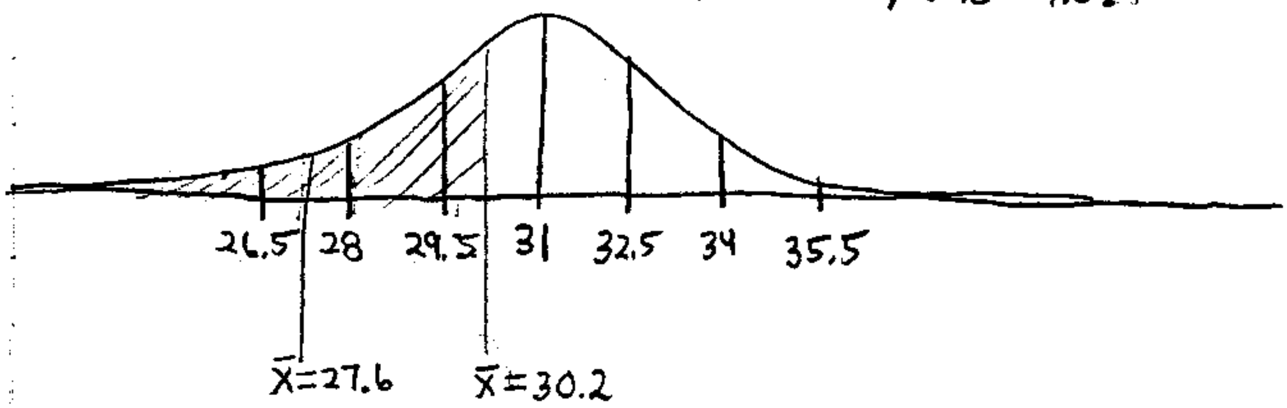
MM: $32.38 \pm 1.96 * 33.37 \sqrt{1 + \frac{1}{201}} = 32.38 \pm 65.57 = (0, 97.95)$

Pharmacies: $48.60 \pm 1.96 * 35.62 \sqrt{1 + \frac{1}{201}} = 48.60 \pm 69.99 = (0, 118.59)$

Assuming scores must be positive.

Lab 2 Soln.

(6.26) (a) $\bar{X}$ is normal by CLT. Also, $\bar{X}$ has mean 31% and std dev $\sigma/\sqrt{n} = 9.6/\sqrt{40} = 1.5$.



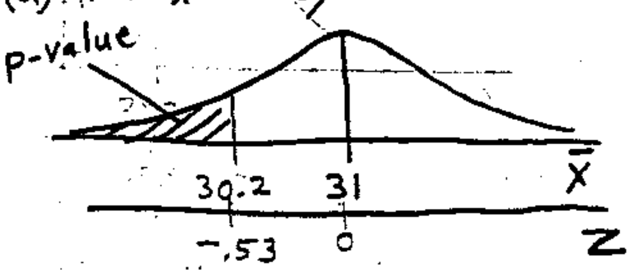
(b) If the mean percentage on household spending was 31, $\bar{X} = 27.6$ is unlikely since it is far out in the tail. On the other hand, $\bar{X} = 30.2$ is likely since it is in the middle of $\bar{X}$'s sampling distribution. Thus, $\bar{X} = 27.6$ suggests H_0 is false, $\bar{X} = 30.2$ does not provide evidence against H_0.

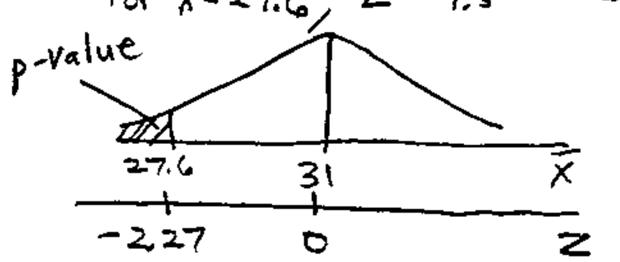
(c) see sketch

(6.28) $H_0: \mu = 52{,}500$
$H_a: \mu > 52{,}500$

(6.30) $H_0: \mu = 2.6$
$H_a: \mu \neq 2.6$

(6.32) (a) For $\bar{X} = 30.2$, $Z = \frac{30.2 - 31}{1.5} = -.53$ For $\bar{X} = 27.6$, $Z = \frac{27.6 - 31}{1.5} = -2.27$





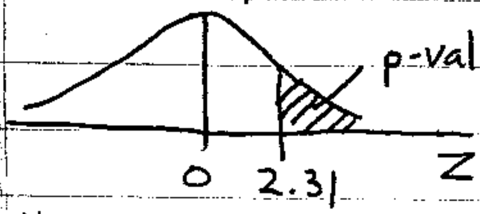
$\boxed{\text{p-value} = 0.2981}$ $\boxed{\text{p-value} = 0.0116}$

(b) Yes. No.

Lab 2 Soln

6.34 (a) $\bar{X} = \dfrac{405 + 378 + 411}{3} = \boxed{398}$

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{398 - 354}{33/\sqrt{3}} = \boxed{2.31}$$

(b) 

p-value $= P(Z > 2.31) = 1 - .9896 = \boxed{.0104}$

(c) Yes. No. Yes.

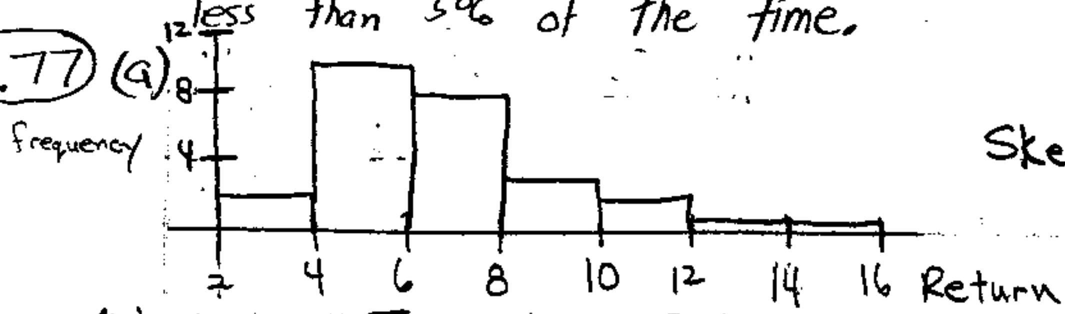
6.52 To say a test is significant at 1% means that if the null hypothesis was in fact true outcomes similar to that obtained would happen less than 1 time per 100 samples. 5% means that if the null hypothesis was true outcomes similar to that obtained would happen less than 5 times per 100 samples.

"Less than 1 time per 100 samples" is "less than 5 times per 100 samples" so 1% significance implies 5% significance. However, the reverse does not hold, so 5% significance does not imply 1% significance.

Lab 2 Soln

(6.53) No. This means that if H_0 is true we observed an outcome that occurs less than 5% of the time.

(6.77) (a)



Skewed Right!

(b) $\bar{X} \pm Z^* \frac{\sigma}{\sqrt{n}}$ $\bar{X} = 7.08$, from JmP.

$7.08 \pm 1.645 \frac{2.75}{\sqrt{27}}$

$7.08 \pm .87$

$(6.21, 7.95)$

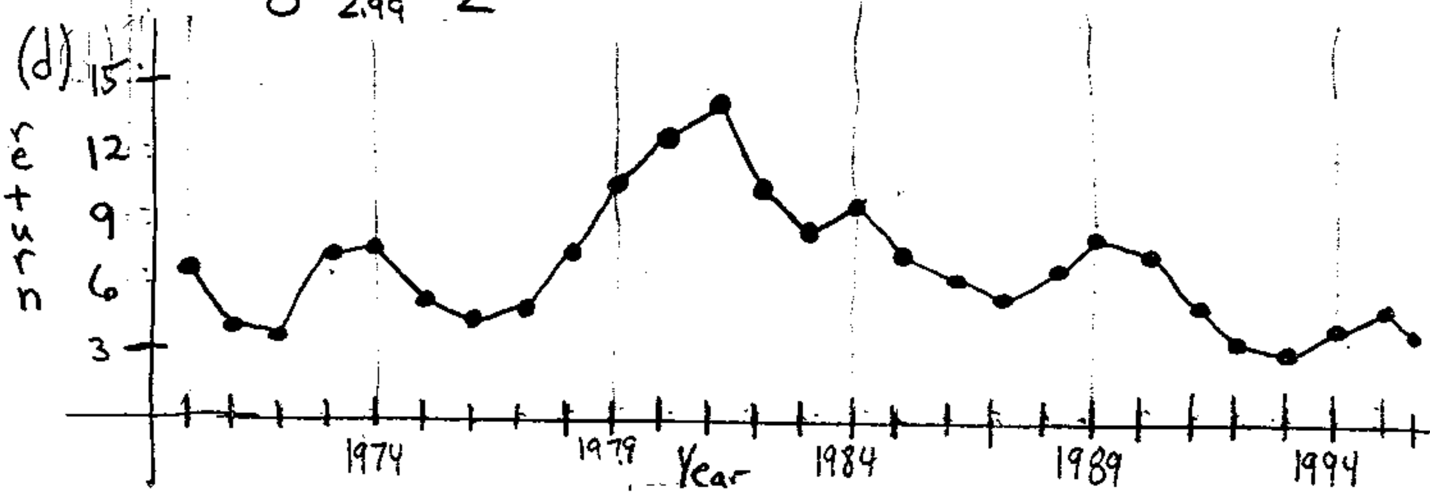
(c) $H_0: \mu = 5.5$

$H_a: \mu > 5.5$

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{7.08 - 5.5}{2.75/\sqrt{27}} = \boxed{2.99}$$

p-value $= P(Z > 2.99) = 1 - .9986 = \boxed{.0014}$

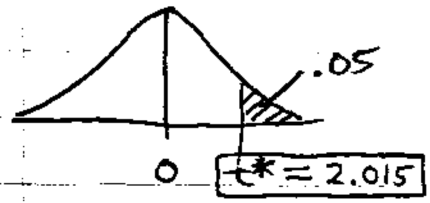
(d)

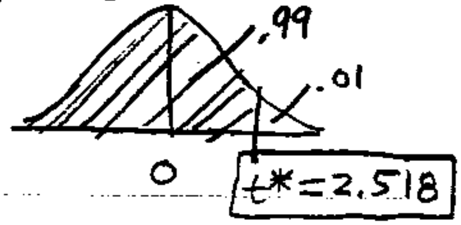


Additional Part for 6.77: Because returns are not normal but skewed right.

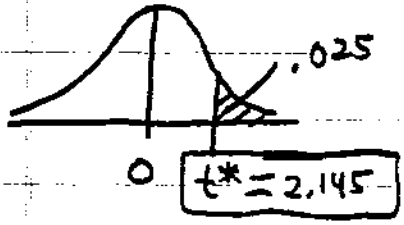
Lab 2 Soln

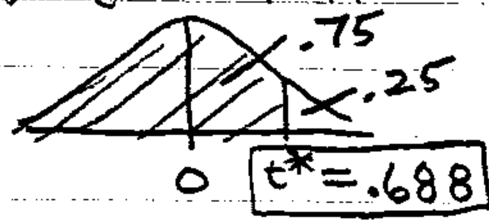
7.2 (a) df = 5 (b) df = 21





7.3 (a) df = n−1 = 14 (b) df = n−1 = 19





7.11 (b)

Subject	D_c Depression (caffeine)	D_p Depression (placebo)	$D_c - D_p$
1	5	16	−11
2	5	23	−18
3	4	5	−1
4	3	7	−4
5	8	14	−6
6	5	24	−19
7	0	6	−6
8	0	3	−3
9	2	15	−13
10	11	12	−1
11	1	0	+1

98	−1	98
31	−1	31
66	−0	66
1341	−0	4311
1	+0	1

Basically normal with a slight skew left.

From JMP, for $D_c - D_p$ column, $\bar{X} = -7.36$, $s = 6.92$.

$H_0: \mu = 0$

$H_a: \mu < 0$

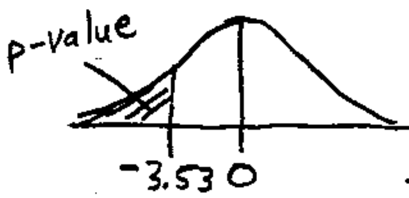
$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{-7.36 - 0}{6.92/\sqrt{11}} = -3.53$$

df = n−1 = 10

df	.005	.0025
10	3.169	3.581

3.53

$\boxed{.0025 < p\text{-value} < .005}$



With a small p-value, yes the data present evidence to suggest that caffeine deprivation increases depression.

Lab 2 Soln

(7.12) (a) $H_0: \mu = 0$

$H_a: \mu > 0$

$t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{332-0}{108/\sqrt{200}} = 43.47$, so p-value ≈ 0 Yes.

(b) $\bar{x} \pm t^* \frac{s}{\sqrt{n}} \approx \bar{x} \pm z^* \frac{s}{\sqrt{n}}$ for large df

$332 \pm 2.576 \frac{108}{\sqrt{200}}$

332 ± 19.7

$(312.3, 351.7)$

(c) The t procedure is OK because

① it is robust in non-normal scenarios as described on pg 379 and

② the credit limit enforced by the bank limits the amount to which the distribution of charges is skewed right (i.e., non normal).

(d) Compare this sample of 200 customers on the no-fee offer to a control group of 200 other customers that are not on the no-fee offer.

(7.15) (a) $\bar{x} \pm t^* \frac{s}{\sqrt{n}} \approx \bar{x} \pm z^* \frac{s}{\sqrt{n}} = 24 \pm 1.96 \frac{11}{\sqrt{75}} = 24 \pm 2.5$

For large n. $= (21.5, 26.5)$

(b) Because the t procedure is robust -- and especially so when n is large. See tan box on page 380.

<u>Additional Part for 7.15</u>: Although the t confidence interval for μ is robust to nonnormality, the t prediction interval for an individual is not, and we are sampling from a skewed-right, non-normal population.

Lab 2 Soln.

More Real Inference

① (a) $\bar{x} \pm t^* \frac{s}{\sqrt{n}}$

$.76 \pm 1.96 \frac{.26}{\sqrt{40}}$

$.76 \pm .08$

$(.68, .84)$

(b) $\bar{x} \pm t^* s\sqrt{1+\frac{1}{n}}$

$.76 \pm 1.96 * .26\sqrt{1+\frac{1}{40}}$

$.76 \pm .52$

$(.24, 1.28)$

Note: $t^* \approx z^*$ for large n

② From JMP, Capital Appreciation $\bar{X} = 1.035$ $s = .07188$ Equity Income Or. $\bar{X} = .6983$ $s = .04875$

(a) $H_0: \mu_{CA} - \mu_{EIO} = 0$

$H_a: \mu_{CA} - \mu_{EIO} \neq 0$

Moore's Method

$$t = \frac{\bar{X}_{CA} - \bar{X}_{EIO} - 0}{\sqrt{\frac{s_{CA}^2}{n_{CA}} + \frac{s_{EIO}^2}{n_{EIO}}}} = \frac{1.035 - .6983}{\sqrt{\frac{.07188^2}{4} + \frac{.04875^2}{6}}}$$

$= 8.20$ df $= 3$ $.001 < \frac{p}{2} < .0025$

$.002 < p < .005$

Dielman's 2nd Method

$$S_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{3(.07188^2) + 5(.04875^2)}{8}$$

$= .003423$

$$t = \frac{\bar{X}_{CA} - \bar{X}_{EIO} - 0}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{1.035 - .6983}{\sqrt{.003423\left(\frac{1}{4} + \frac{1}{6}\right)}}$$

$= 8.92$ df $= 8$ $\frac{p}{2} < .0005$

p-value $< .001$

Both methods suggest that the mean expense ratio for the 2 types of funds are not the same.

(b) $\bar{X}_1 - \bar{X}_2 \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

$1.035 - .6983 \pm 3.182\sqrt{\frac{.07188^2}{4} + \frac{.04875^2}{6}}$

$.3367 \pm .1307$

$(.2060, .4674)$

$\bar{X}_1 - \bar{X}_2 \pm t^* \sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

$1.035 - .6983 \pm 2.306\sqrt{.003423\left(\frac{1}{4} + \frac{1}{6}\right)}$

$.3367 \pm .0871$

$(.2496, .4238)$