

Sample problems—Chapter 5

1. Find, assuming Z is a standard normal random variable:
 - a. $P(Z > 1.5)$
 - b. $P(Z < -7)$
 - c. $P(-.25 < Z < 1.25)$
 - d. $P(Z < 2.62)$
2. Z is standard normal. Find C such that:
 - a. $P(Z > c) = .02$
 - b. $P(Z > c) = .72$
 - c. $P(Z < c) = .40$
3. Suppose 90% of overseas calls are completed. A company makes 100 calls. What is the probability that more than 95% of the calls go through?
4. In a study of London bus drivers, measurements of daily food intake in calories were taken. The drivers had a mean intake of 2821 calories with a sd of 436 calories. The distribution was approximately normal. What is the probability that a randomly selected bus driver will eat more than 3000 calories today?
5. In Hometown, Michigan, 65% of households have an annual income greater than \$14,000. Assume a sample of 5 households has been selected randomly. What is the probability that at least 4 of these households have an income greater than \$14,000?
6. Suppose that 6% of all forest tracts were damaged last summer in fires in Oregon. If the Department of Natural Resources owns 1000 such tracts, what is the probability that less than 5% of these were damaged by fire last summer?
7. A basketball player makes 75% of her free throws over the course of a season. What's the probability that in a key game, she misses 5 out of 12 free throws?
8. Assume that the length of your telephone conversations is distributed normally with an average length of 4.5 minutes and a variance of 4.0.
 - a. What is the probability that the next call you make will last longer than 5 minutes?
 - b. Suppose you will make 20 calls in one week. What is the probability that the average length of that week's calls will be less than 3.0 minutes?

9. According to government data, 22% of American children under the age of six live in households with incomes lower than the official poverty level. A random sample of 300 children is selected for a study in learning in early childhood. Calculate the probability that at least 80 of the children in the sample live in poverty.
10. The time that a technician requires to perform preventive maintenance operations on an air conditioning unit averages one hour with a sd of 1 hour. Your company operates 70 of these units. What is the probability that their average maintenance time exceeds 50 minutes?
11. A manufacturer claims that at least 95% of the items he produces are failure-free. Examination of a random sample of 600 items showed 39 to be defective. Does this evidence support his claim?
12. An environmental study has shown that the daily average noise level on a busy street is 37 decibels with sd of 6 decibels. Decibel levels were recorded on 30 random busy streets in each of 20 cities (so, 600 streets). Then mean noise level per sample was calculated for each street. What is the decibel level below which only 2% of my sample means should fall?
13. An insurance company has long term records indicating that the probability that a married male 30 years of age will survive to at least age 60 is .66. Of a given set of 100 such males, what is the probability that at least 75 survive to age 60?
14. Suppose that the weight of a certain breed of rabbit is normally distributed with a mean of 10 lbs and a variance of 8.
- Above what weight do 75% of the rabbits weigh?
 - What is the range of the middle 90% of weights?
 - What is the probability of a rabbit weighing within 2 lbs of the mean?
15. Suppose that in the long run annual precipitation in Illinois cities is $N(33.18 \text{ in}, 4.23 \text{ in})$.
- What is the probability that next year, annual precipitation will be 40 inches or more?
 - What is the probability that the mean rainfall over the next 2 years will be 40 inches or more?

Answers to extra practice problems - Ch 5

- 1 a. $P(Z > 1.5) = .0668$
 b. .2420
 c. $.8944 - .4013 = .4931$
 d. .9956

- 2 a. $P(Z > c) = .02$ $P(Z \leq c) = .98$ $c = 2.05$
 b. $P(Z > c) = .72$ $P(Z \leq c) = .28$ $c = -.585$
 c. $P(Z < c) = .40$ $c = -.255$

- 3 $n = 100$ $p = .90 \rightarrow$ (binomial) mean = $\frac{np}{n}$ $sd = \sqrt{\frac{np(1-p)}{n}}$ $\left. \begin{matrix} np \geq 10 \\ n(1-p) \geq 10 \end{matrix} \right\}$ use normal approx.
 $P(\hat{p} > .95) = P\left(Z > \frac{.95 - .90}{\sqrt{\frac{.90 \times .10}{100}}}\right) = P\left(Z > \frac{.05}{.03}\right) = P(Z > 1.67) = .0475$

- 4 $\mu = 2821$ $\sigma = 436 \rightarrow$ (quantitative)
 $P(X > 3000) = P\left(Z > \frac{3000 - 2821}{436}\right) = P(Z > .41) = .3409$

- 5 $n = 5$ $p = .65$ ^(binomial) Need Table C.
 Since $p > .50$ (and Table C only goes up to .50), we need to think "backwards."
 $p = .65 =$ prob. of annual income greater than \$14,000.
 $P_{\text{Backwards}} = .35 =$ prob. of " " less than " "

$P(\text{at least 4 with incomes} > \$14,000) = P(4) + P(5)$

If 4 have high incomes, 1 does not.

If 5 have high incomes, 0 do not.

We want $P(0 \text{ or } 1 \text{ with incomes} < 14,000) = P(0) + P(1)$
 when $p_B = .35$

From Table C, $n = 5$, $p_B = .35$ $P(0) + P(1) = .1160 + .312$
= .4284

6) $p = .06$ (binomial) $\left. \begin{array}{l} np \geq 10 \\ n(1-p) \geq 10 \end{array} \right\}$ so use normal approx.

$$P(p < .05) = P\left(Z < \frac{.05 - .06}{\sqrt{\frac{.06 \times .94}{1000}}}\right) = P\left(Z < \frac{-.01}{.00751}\right) = P(Z < -1.33) = \boxed{.0918}$$

7) $X = \# \text{ missed} = 5$ prob. she makes = .75, so prob. she misses = .25
 (binomial) $\left. \begin{array}{l} p = .25 \\ n = 12 \end{array} \right\}$ $np < 10$ so use Table C. When $n = 12, p = .25$
 $P(X = 5) = \boxed{.1032}$

8) mean = 4.5, variance = 4, sd = 2 (quantitative)

a. $P(X > 5) = P\left(Z > \frac{5 - 4.5}{2}\right) = P(Z > .25) = .4013$

b. $P(\bar{X} < 3) = P\left(Z < \frac{3 - 4.5}{2/\sqrt{50}}\right) = P\left(Z < \frac{-1.5}{.283}\right) = P(Z < -5.33) = \boxed{.0004}$

9) $p = .22$ $\left. \begin{array}{l} np \geq 10 \\ n(1-p) \geq 10 \end{array} \right\}$ so use normal approx. $\left\{ \begin{array}{l} \text{mean} = np = 66 \\ \text{sd} = \sqrt{np(1-p)} = 7.17 \end{array} \right.$

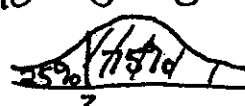
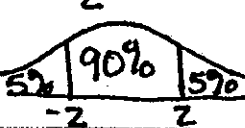
$$P(X \geq 80) = P\left(Z \geq \frac{80 - 66}{7.17}\right) = P(Z \geq 1.95) = \boxed{.0256}$$

10) mean = 60, sd = 60, $n = 70$
 $P(\bar{X} > 50) = P\left(Z > \frac{50 - 60}{60/\sqrt{70}}\right) = P\left(Z > \frac{-10}{7.17}\right) = P(Z > -1.39) = \boxed{.9177}$

11) $p = .95$ $\left. \begin{array}{l} np = 570 = \text{mean} \\ n(1-p) = 5.34 = \text{sd} \end{array} \right\}$ $\left. \begin{array}{l} \text{mean} = 570 \\ \text{sd} = 5.34 \end{array} \right\}$ $\left. \begin{array}{l} \text{mean} = np = 66 \\ \text{sd} = \sqrt{np(1-p)} = 7.17 \end{array} \right\}$ avg. # defects = 600(.05) = 30
 $P(X \geq 39) = P\left(Z \geq \frac{39 - 30}{5.34}\right) = P(Z \geq 1.69) = \boxed{.0455}$ This evidence does not support his claim

12) $P(X < Z) = .02$ $Z = -2.05 = \frac{L - 37}{6/\sqrt{600}}$ $L = 36.5 \text{ decibels}$

13) $p = .66$ $\left. \begin{array}{l} \text{mean} = np = 66 \\ \text{sd} = \sqrt{np(1-p)} = 4.74 \end{array} \right\}$
 $P(X \geq 75) = P\left(Z \geq \frac{75 - 66}{4.74}\right) = P(Z \geq 1.90) = \boxed{.0287}$

14) $\mu = 10$ $\sigma^2 = 8$ $\sigma = 2.83$
 a.  $Z = -0.675 = \frac{W - 10}{2.83}$ $W = 8.09 \text{ lb.}$ $\boxed{75\% \text{ of rabbits weigh } > 8.09}$
 b.  $Z = 1.645$ $-1.645 = \frac{W - 10}{2.83}$ $W = 5.34$ on lower tail
 $1.645 = \frac{W - 10}{2.83}$ $W = 14.66$ on upper tail

The middle 90% of weights range from 5.34 to 14.66 $- 5.34 = 9.32$

c. $P(8 < X < 12) = P\left(\frac{8 - 10}{2.83} < Z < \frac{12 - 10}{2.83}\right) = P(-.71 < Z < .71) = .76$

15) a. $P(X \geq 40) = P\left(Z \geq \frac{40 - 33.18}{4.23}\right) = P(Z \geq 1.61) = \boxed{.0537}$
 b. $P(\bar{X} \geq 40) = P\left(Z \geq \frac{40 - 33.18}{4.23/\sqrt{19}}\right) = P(Z \geq 2.28) = \boxed{.0113}$

Sample problems (ch. 6, 7)

1. Suppose the emission of carbon monoxide (CO) was measured from 22 cars. The mean and standard deviation are 7.48 and 4.35, respectively. Compute a 95% confidence interval for the mean amount of CO emitted from cars.
2. A professor would like to estimate the difference in mean scores for students taking an exam at 8 am and students taking the same exam at 2 pm. A random sample of students produces the following data:

<u>Time</u>	<u>n</u>	<u>\bar{x}</u>	<u>s</u>
8 am	42	72.4	10.4
2 pm	38	76.3	5.1

- a. Find a 99% confidence interval for the difference in means.
 - b. Is there a significant difference in means ($\alpha = .05$)?
3. A training director wishes to find out whether or not a training program will increase employee efficiency. He takes a random sample of 10 employees and measures their efficiency before and after the training program. The efficiency rating increased, on average, by 4.6 with a standard deviation of 4.40.
 - a. Is this increase statistically significant ($\alpha = .10$)?
 - b. Compute a 95% confidence interval for the average change in efficiency.
 4. A football coach said that, based on past records, the mean weight of the defensive linemen is 235 pounds. A sample of 10 defensive linemen this year revealed that the mean weight is 240 pounds, and the sd of the sample is 11 pounds. At the 0.01 level, is this sufficient evidence that the mean weight has increased?
 5. A manufacturer wants to estimate the average lifetime (in hours) of his batteries. He knows that battery lifetime is normally distributed with a standard deviation of 10. If he wants to be 95% confident that his sample mean is no more than 2.5 hours off, how big a sample does he need?
 6. The melting point of a certain alloy is required to be 1000°C to meet certain specifications. A manufacturer tested 30 samples of the alloy and found that the average melting point was 975°C, with a standard deviation of 40°C. If the melting point is significantly lower than 1000°C at the .05 level of significance, the manufacturer must reject the whole batch.
 - a. Must she reject the batch?
 - b. Compute a 95% confidence interval for the average melting point.

7. The length of time it takes to complete production of a part was measured for both an old and a new method in two independent random samples.

Method	n	\bar{x}	s^2
Old	16	53.1	327.61
New	12	40.7	77.44

- a. Does the new method take less time ($\alpha = .10$)?
 b. Compute a 95% confidence interval for the difference in time between the two methods.

8. To test the effectiveness of a diet, a random sample of six individuals are weighed before starting the diet and then again after 3 months on the diet. The data are as follows:

Individual	Before	After
1	163	142
2	212	183
3	195	196
4	184	153
5	168	136
6	195	186

$$\bar{D} = 20.17$$

$$s_D = 13.48$$

- a. Was the diet effective? (significance level=.01)
 b. Compute a 95% confidence interval for the difference between their before and after weights.

9. An advertising firm plans a survey to estimate the average amount of weekend time adults spend watching television. If viewing time is a normal random variable with a standard deviation of 2 hours, how large a sample is needed for 99% confidence that the sample mean is within 15 minutes of the population mean?

Chapter 6,7 formulas:

Sample size

$$n \geq [z_{\alpha/2} \sigma / m]^2 \quad \text{where } z_{\alpha/2} = \begin{matrix} 90\% & 95\% & 99\% \\ 1.645 & 1.96 & 2.576 \end{matrix}$$

One-sample

Hypothesis testing for a sample mean: $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ $\rightarrow n-1$ degrees of freedom.

CI for a sample mean: $\bar{x} \pm t_{\alpha/2, n-1} s/\sqrt{n}$

The margin of error: $t_{\alpha/2, n-1} s/\sqrt{n}$

Paired Samples

Hypothesis testing for $(\mu_1 - \mu_2)$, paired samples: $t = \frac{\overline{\text{Diff}} - \mu_{\text{Diff}}}{s_{\text{Diff}}/\sqrt{n}}$ $\rightarrow df = \# \text{pairs} - 1$
 where $n = \# \text{pairs}$

$(1-\alpha)$ confidence interval for $(\mu_1 - \mu_2)$: paired samples: $\overline{\text{Diff}} \pm t_{n-1, \alpha/2} s_{\text{Diff}}/\sqrt{n}$ $\rightarrow df = \# \text{pairs} - 1$

Two-samples: Variances unequal

Hypothesis testing: $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(s_1^2/n_1 + s_2^2/n_2)}}$ $\rightarrow df = \text{smaller of } n_1 - 1 \text{ and } n_2 - 1.$

$1-\alpha$ CI for $(\mu_1 - \mu_2)$: $(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{(s_1^2/n_1 + s_2^2/n_2)}$ $\rightarrow df = \text{smaller of } n_1 - 1 \text{ and } n_2 - 1.$

Answers to sample problems (Ch. 6, 7)

1) $\bar{x} \pm t_{n-1, 95\%} s/\sqrt{n}$

$t_{21 \text{ df}, 95\%} = 2.080$

$7.48 \pm 2.080(4.35)/\sqrt{22}$

7.48 ± 1.93

$(5.55, 9.41)$

2) a) 99% CI $(\bar{x}_1 - \bar{x}_2) \pm t \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ $df = 38 - 1 = 37$

$t_{30 \text{ df}, 99\%} = 2.75$

$(72.4 - 76.3) \pm 2.75 \sqrt{\frac{(10.4)^2}{42} + \frac{(5.1)^2}{38}}$

-3.9 ± 4.95

$(-8.85, 1.05)$

b) $H_0: \mu_{8:00} = \mu_{2:00}$

$H_a: \mu_{8:00} \neq \mu_{2:00}$

or

$H_0: \mu_{8:00} - \mu_{2:00} = 0$

$H_a: \mu_{8:00} - \mu_{2:00} \neq 0$

$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{72.4 - 76.3}{1.8} = -2.17$ $df = 37$

Look at $df = 30$ in the table. $\Pr(t < -2.17)$ is between .01 & .02. To get the p-value, we need to double this (2 sided H_a) so p is between .02 and .04. $p < .05$ so reject H_0 . There is a signif. diff. in med.

3) Paired t-test.

a) $H_0: \mu_{diff} = 0$

$H_a: \mu_{diff} > 0$

$t = \frac{4.6 - 0}{4.4/\sqrt{10}} = 3.31$ $df = 10 - 1 = 9$

p-value = $\Pr(t > 3.31)$ = between .0025 and .005. $p < \alpha$, reject H_0 . The increase is stat. signif.

b) CI. $Diff \pm t_{n-1, 95\%} s_{diff}/\sqrt{n}$

$t_{9 \text{ df}, 95\%} = 2.262$

$4.6 \pm 2.262(4.4)/\sqrt{10}$
 4.6 ± 3.15

$(1.45, 7.75)$

4) $H_0: \mu = 235$

$H_a: \mu > 235$

$t = \frac{240 - 235}{11/\sqrt{10}} = 1.44$ $df = 9$

p-value = $\Pr(t > 1.44)$
p is between .05 and .10
 $p > \alpha$, $p > .01$, no evidence weight has increased.

$$\boxed{5} \quad n \geq \left[\frac{z_{\alpha/2} \sigma}{m} \right]^2 = \left[\frac{(1.96)(10)}{2.5} \right]^2 = 61.5 \quad \boxed{n=62}$$

$$\boxed{6} \quad a) \quad H_0: \mu = 1000 \quad t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \quad df = n - 1$$

$$H_a: \mu < 1000$$

$$t = \frac{975 - 1000}{40/\sqrt{30}} = -3.4 \quad df = 29$$

p-value = $\Pr(t < -3.4)$ = between .001 and .0005.
 $p < \alpha$ so reject H_0 . She must reject the batch

$$b) \quad \bar{x} \pm t_{\alpha/2, n-1} s/\sqrt{n} \quad t_{.95, df=29} = 2.045$$

$$975 \pm 2.045(40/\sqrt{30}) \quad 975 \pm 14.9 \quad \boxed{(960.1, 989.9)}$$

$$\boxed{7} \quad a) \quad H_0: \mu_{OLD} = \mu_{NEW} \quad t = \frac{53.1 - 40.7}{\sqrt{\frac{327.61}{16} + \frac{77.44}{12}}} = \frac{12.4}{5.19} = 2.39$$

$$H_a: \mu_{OLD} > \mu_{NEW}$$

$df = 12 - 1 = 11$. $\Pr(t > 2.39)$ = p-value = between .01 and .02
 $p < .10$. $p < \alpha$. reject H_0 . The new method takes less t

$$b) \quad 95\% \text{ CI, } (53.1 - 40.7) \pm t_{11, 95\%} \sqrt{\frac{327.61}{16} + \frac{77.44}{12}}$$

$$t_{11, 95\%} = 2.201 \quad 12.4 \pm 2.201(5.19) \quad \boxed{(9.8, 23.02)}$$

$$\boxed{8} \quad a) \quad H_0: \mu_{\text{before-after}} = 0 \quad t = \frac{\overline{\text{Diff}} - \mu_{\text{diff}}}{s_{\text{diff}}/\sqrt{n}} = \frac{20.17 - 0}{13.48/\sqrt{6}} = 3.67$$

$$H_a: \mu_{\text{before-after}} > 0$$

$$df = 5$$

$\Pr(t > 3.67)$ = between .005 and .01. $p < .01$ so
 reject H_0 and conclude the diet is effective

$$b) \quad \overline{\text{Diff}} \pm t_{n-1, 95\%} s_{\text{diff}}/\sqrt{n} \quad t_{5, 95\%} = 2.571$$

$$20.17 \pm 2.571(13.48)/\sqrt{6} \quad 20.17 \pm 14.15$$

$$\boxed{(6.02, 34.32)}$$

$$\boxed{9} \quad n \geq \left[\frac{z_{\alpha/2} \sigma}{m} \right]^2 = \left[\frac{(2.576)(2)}{(.25)} \right]^2 = 424.7$$

$$\boxed{n=425}$$

Sample problems (ch. 8)

1. In a random telephone sample of 500 homes, 450 of them had the t.v. on. Find a 95% confidence interval for the proportion of viewers in the population.
2. A sample of 500 people were classified as being athletic or nonathletic. Among 200 classified as athletic, it was found that 52 regularly eat Wheaties. Among the 300 nonathletic persons, 48 ate Wheaties. Find a 99% confidence interval for the difference in the proportions that eat Wheaties.
3. The 1958 Detroit Area Study investigated the influence of religion on everyday life. It contained a survey of the population of the metropolitan area. Of the 656 respondents, 267 were Protestants and 230 were Catholics. One question asked whether the government was doing enough in areas such as housing, unemployment and education. 161 of the Protestants and 136 of the Catholics said, "No." Is there evidence that Protestants and Catholics differed on this issue ($\alpha=.01$)?
4. A bank has three branch offices. A random sample of 200 customers showed 75 who banked at the office on 30th Street. Do these data indicate that more than 1/3 of the customers prefer to use this office?
5. Overall, 90% of students at Zwingli College attend class regularly. A teacher there would like to estimate the proportion of her students who attend class on a regular basis. How large a sample should be taken if she would like to be 99% confident that her estimate is in error by no more than 4 percentage points (.04)?

6. A good fielder whose batting average is only 0.150 gets glasses. In his next 50 at bats he gets 16 hits (batting average = $16/50 = .320$). Did the glasses help? Use $\alpha = .05$. Then compute a 90% confidence interval for his recent batting average (.320).
7. A land developer in Florida wants to know what percent of people who plan to purchase a vacation home in Florida would like to buy a condominium. She wants to be 95% confident that she is within 5 percentage points of the true proportion. How many people must she include in her sample?
8. A mock model of a proposed new automobile was shown to two groups of 150 each. One group consisted of people between 18 and 25 years of age, and the other group consisted of persons more than 50 years old. 80% of the younger group rated the styling satisfactory, but only 50% of the older group gave it a similar rating. In evaluating the market potential of this proposed automobile, is it reasonable to expect that it would appeal primarily to younger persons? Or is it possible that this difference of 30 percentage points could be due to sampling, and might not the two age groups in the population like the proposed automobile equally well?
- Next, compute a 95% confidence interval for the difference in proportions that rated the automobile design satisfactory.

Sample problems (ch. 8)

1. In a random telephone sample of 500 homes, 450 of them had the t.v. on. Find a 95% confidence interval for the proportion of viewers in the population.

$$\hat{p} = \frac{450}{500} = .90 \quad CI: \hat{p} \pm Z \sqrt{\hat{p}(1-\hat{p})/n}$$

$$.90 \pm 1.96 \sqrt{(.90)(.10)/500}$$

$$.90 \pm .026$$

(.874, .926)

2. A sample of 500 people were classified as being athletic or nonathletic. Among 200 classified as athletic, it was found that 52 regularly eat Wheaties. Among the 300 nonathletic persons, 48 ate Wheaties. Find a 99% confidence interval for the difference in the proportions that eat Wheaties.

$$\hat{p}_{\text{athletic}} = 52/200 = .26 \quad \hat{p}_{\text{nonathletic}} = 48/300 = .16$$

$$(\hat{p}_1 - \hat{p}_2) \pm Z \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$(.26 - .16) \pm 2.576 \sqrt{\frac{(.26)(.74)}{200} + \frac{(.16)(.84)}{300}}$$

$$.10 \pm 2.576 (.038) \quad .10 \pm .098$$

(.002, .198)

3. The 1958 Detroit Area Study investigated the influence of religion on everyday life. It contained a survey of the population of the metropolitan area. Of the 656 respondents, 267 were Protestants and 230 were Catholics. One question asked whether the government was doing enough in areas such as housing, unemployment and education. 161 of the Protestants and 136 of the Catholics said, "No." Is there evidence that Protestants and Catholics differed on this issue ($\alpha = .01$)?

$$H_0: p_1 = p_2 \quad H_a: p_1 \neq p_2$$

$$p_p = 161/267 = .603 \quad p_c = 136/230 = .591$$

$$\hat{p} = \frac{161 + 136}{267 + 230} = .598$$

$$Z = \frac{p_p - p_c}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{.603 - .591}{\sqrt{(.598)(.402)\left(\frac{1}{267} + \frac{1}{230}\right)}} = \frac{.012}{.044} = .273$$

$p(27,27) = .39$
 2 sided
 $p = 2(.39) = .78$
 accept H_0

4. A bank has three branch offices. A random sample of 200 customers showed 75 who banked at the office on 30th Street. Do these data indicate that more than 1/3 of the customers prefer to use this office?

$$\hat{p} = 75/200 = .375$$

$$H_0: p = 1/3 \quad H_a: p > 1/3$$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.375 - .33}{\sqrt{\frac{(.33)(.67)}{200}}} = \frac{.045}{.033} = 1.36$$

$P_r(Z > 1.36) = 1 - .9131 = .0869$
 accept H_a

more than 1/3 of the customers do not prefer to use this office

5. Overall, 90% of students at Zwingli College attend class regularly. A teacher there would like to estimate the proportion of her students who attend class on a regular basis. How large a sample should be taken if she would like to be 99% confident that her estimate is in error by no more than 4 percentage points (.04)?

$$n = \left[\frac{Z_{\alpha/2}}{m} \right]^2 p(1-p) = \left[\frac{2.576}{.04} \right]^2 (.9)(.1) = 373.3$$

n = 374

6. A good fielder whose batting average is only 0.150 gets glasses. In his next 50 at bats he gets 16 hits (batting average = $16/50 = .320$). Did the glasses help? Use $\alpha = .05$. Then compute a 90% confidence interval for his recent batting average (.320).

$$H_0: p = .150$$

$$H_a: p > .150$$

$$z = \frac{.32 - .15}{\sqrt{\frac{(.15)(.85)}{50}}} = \frac{.17}{.05} = 3.4$$

$$P(Z > 3.4) = 1 - .9997 = .0003$$

$p < .05$, reject H_0 , The glasses help

confidence interval $\hat{p} \pm z_{\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n}$

$$.320 \pm 1.645 \sqrt{(.32)(.68)/50}$$

$$.32 \pm .109$$

$$(.211, .429)$$

7. A land developer in Florida wants to know what percent of people who plan to purchase a vacation home in Florida would like to buy a condominium. She wants to be 95% confident that she is within 5 percentage points of the true proportion. How many people must she include in her sample?

$$n \geq \left[\frac{z_{\alpha/2}}{m} \right]^2 p(1-p) = \left[\frac{1.96}{.05} \right]^2 (.5)(.5) = 384.2$$

← since no prior estimate of

$$n = 385$$

8. A mock model of a proposed new automobile was shown to two groups of 150 each. One group consisted of people between 18 and 25 years of age, and the other group consisted of persons more than 50 years old. 80% of the younger group rated the styling satisfactory, but only 50% of the older group gave it a similar rating. In evaluating the market potential of this proposed automobile, is it reasonable to expect that it would appeal primarily to younger persons? Or is it possible that this difference of 30 percentage points could be due to sampling, and might not the two age groups in the population like the proposed automobile equally well?

$$\hat{p}_1 = .80$$

$$\hat{p}_2 = .50$$

$$n_1 = 150$$

$$n_2 = 150$$

#younger that like the car = $(.8)(150) = 120$ #older = $(.50)(150) = 75$

Next, compute a 95% confidence interval for the difference in proportions that rated the automobile design satisfactory.

$$H_0: p_{\text{younger}} = p_{\text{older}}$$

$$H_a: p_y > p_o$$

$$\hat{p} = \frac{120 + 75}{300} = .65$$

$$z = \frac{.80 - .50}{\sqrt{(.65)(.35)\left(\frac{1}{150} + \frac{1}{150}\right)}} = \frac{.3}{.055} = 5.45$$

$$P(Z > 5.45) < .0001, p < \alpha, \text{Reject } H_0$$

This design appeals more to the younger

confidence interval: $(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

$$(.8 - .5) \pm 1.96 \sqrt{\frac{(.8)(.2)}{150} + \frac{(.5)(.5)}{150}}$$

$$.3 \pm .1$$

$$(.2, .4)$$

we are 95% confident that the true diff. in props. is b/w 20%

Ch. 9 #13

	Rows: This year		Columns: Next year	
	Loser	Winner	Loser	Winner
This year Loser	69.17	30.83	100.00	
	70.34	30.33	50.00	
	83	37	120	
	59.00	61.00	120.00	
This year Winner	29.17	70.83	100.00	
	29.66	69.67	50.00	
	35	85	120	
	59.00	61.00	120.00	
All	49.17	50.83	100.00	
	100.00	100.00	100.00	
	118	122	240	
	118.00	122.00	240.00	

Chi-Square = 38.411, DF = 1, P-Value = 0.000

Cell Contents --
 % of Row
 % of Col
 Count
 Exp Freq

Ch 9 #17

	Rows: alcohol			Columns: nicotine	
	1. none	2. 1-15	3. 16+	All	
1. none	85.37	5.69	8.94	100.00	
	34.54	10.77	13.25	27.21	
	105	7	11	123	
	82.73	17.69	22.59	123.00	
2. low	76.32	6.58	17.11	100.00	
	19.08	7.69	15.66	16.81	
	58	5	13	76	
	51.12	10.93	13.96	76.00	
3. med	51.53	22.70	25.77	100.00	
	27.63	56.92	50.60	36.06	
	84	37	42	163	
	109.63	23.44	29.93	163.00	
4. high	63.33	17.78	18.89	100.00	
	18.75	24.62	20.48	19.91	
	57	16	17	90	
	60.53	12.94	16.53	90.00	
All	67.26	14.38	18.36	100.00	
	100.00	100.00	100.00	100.00	
	304	65	83	452	
	304.00	65.00	83.00	452.00	

Chi-Square = 42.252, DF = 6, P-Value = 0.000

Cell Contents --
 % of Row
 % of Col
 Count
 Exp Freq

- what are the null and alternate hypotheses
- Describe any relations between these two categorical variable using conditional perc
- what is the actual who were winners both years?
- What is the expected that were winners both years (assuming no asso
- Is the assoc. statistic significant (give the df, and p-value to back up your answer)?

- what are the null and alternate hypotheses?
- Do you think there is assoc. between alcohol & nicotine use (give condi distrib. to back up your answer)?
- Is the assoc. statist significant?

Ch 9 #23

Rows: Group	Columns: Country				
	Ireland	Italy	Norway	Portugal	All
Nontaste	40.98	24.41	14.75	19.85	100.00
	28.74	25.00	30.45	24.01	26.92
	225	134	81	109	549
	210.82	144.32	71.62	122.24	549.00
Tasters	37.45	26.98	12.42	23.15	100.00
	71.26	75.00	69.55	75.99	73.08
	558	402	185	345	1490
	572.18	391.68	194.38	331.76	1490.00
All	38.40	26.29	13.05	22.27	100.00
	100.00	100.00	100.00	100.00	100.00
	783	536	266	454	2039
	783.00	536.00	266.00	454.00	2039.00

Chi-Square = 5.957, DF = 3, P-Value = 0.114

Cell Contents --
 % of Row
 % of Col
 Count
 Exp Freq

a) what are the null and alternate hypotheses?

b) Of the 783 people from Ireland, what percent were tasters?

c) Is the % tasters the same in each country, or does it differ significantly?

One more:

A sample of the employees at a large corporation were asked their opinion regarding a new pension plan. The results are in the table below. Is there a relationship between age and their opinion of the new plan? Test at the 0.05 level.

Crosstabs						
		Age				
Count		20-29	30-39	40-49	50+	
Opinion	Total %					
	Row %					
	Col %					
	Expected					
	Bad	81	41	27	8	157
		18.71	9.47	6.24	1.85	36.26
		51.59	26.11	17.20	5.10	
		50.62	47.13	30.68	8.16	
		58.0139	31.545	31.9076	35.5335	
	Fair	52	29	21	9	111
	12.01	6.70	4.85	2.08	25.64	
	46.85	26.13	18.92	8.11		
	32.50	33.33	23.86	9.18		
	41.0162	22.3025	22.5589	25.1224		
Good	27	17	40	81	165	
	6.24	3.93	9.24	18.71	38.11	
	16.36	10.30	24.24	49.09		
	16.88	19.54	45.45	82.65		
	60.97	33.1524	33.5335	37.3441		
	160	87	88	98	433	
	36.95	20.09	20.32	22.63		

Tests				
Source	DF	-LogLikelihood	RSquare	(U)
Model	6	66.97971	0.1426	
Error	4	402.58028		
C Total	10	469.55999		
Total Count	433			
Test	ChiSquare	Prob>ChiSq		
Likelihood Ratio	133.959	<.0001		
Pearson	128.516	<.0001		

Answers to Extra Ch. 9 problems

Ch. 9 #13.

a) H_0 : There is no assoc. between a fund's performance this year and next year.
 H_a : There is an assoc.

b) We want to see if this year's performance (row variable) predicts next year's performance. Therefore use row %.

This year Loser

Next year loser 69.17%

Next year winner 30.83%

This year Winner

Next year loser 29.17%

Next year winner 70.83%

There appears to be a strong relationship. Only 30.83% of this year's losers are next year winners, while 70.83% of this year's winners are also winners's next year.

c) 85

d) 61

e) Yes ($\chi^2 = 38.411$, $df = 1$, $p < .001$), we reject H_0 .

Ch. 9 #17

a) H_0 : There is no assoc. between alcohol use and nicotine use.

H_a : There is an assoc.

b) There is no explanatory (predictor) variable, so it doesn't matter if we use row % or col %.

Ch. 9 #17 (continued)

b) (cont.) I'll just pick col 9.

of those who use no nicotine:

34.54% drink no alcohol
 19.08% drink low amounts of alc.
 27.63% " med " " "
 18.75% " high " " "

Of those who get 1-15 mg nicotine/day:

10.77% drink no alc.
 7.69% " low "
 56.92% " med. "
 24.62% " high "

Of those who get 16+ mg nicotine/day:

13.25% drink no alc.
 15.66% " low "
 50.60 " med "
 20.48 " high "

There appears to be an assoc. Those who do not smoke are much more likely to avoid alcohol.

c) Yes. ($\chi^2 = 42.252$, $df = 6$, $p < .001$) $p < \alpha$ so reject H_0

Ch. 9 #23

a) H_0 : no assoc. between country and whether or not you are a taster
 H_a : there is an assoc.

b) 71.26%

c) the same. $\chi^2 = 5.957$, $df = 3$, $p = .114$ $p > \alpha$ so accept the null hyp. of no assoc.

One more!

There is a relationship. $\chi^2 = 128.516$, $df = 6$, $p < .0001$. $p < \alpha$ so conclude there is a signif. association.

Let's look at col. 9 to find the nature of the relationship. Age

20-29	30-39	40-49	50+
50.62% bad	47.13% bad	30.68% bad	8.16% bad
32.5% fair	33.33% fair	23.86% fair	9.18% fair
16.88% good	19.54% good	45.45% good	82.65% good

The older you get, the more likely you are to think the pension plan is good.