Answers to additional paper and pencil Questions

Please bring any errors to my attention
Fall 2004

Chapter 4

Question 1

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

a) $.6 = .5 + P(B) - 0$

$.1 = P(B)$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

2) $.6 = .5 + P(B) - .5P(B)$

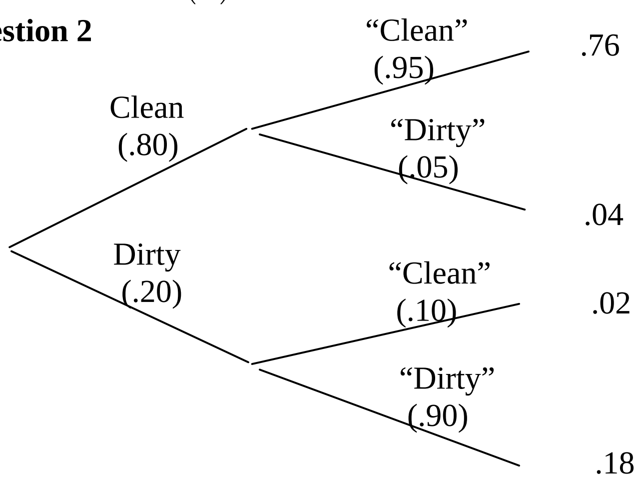
$.2 = P(B)$

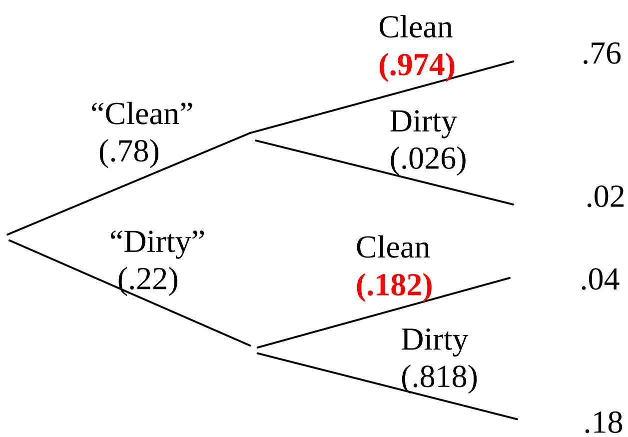
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

3) $.6 = .5 + P(B) - .4P(B)$

$.16666 = P(B)$

Question 2





Question 3

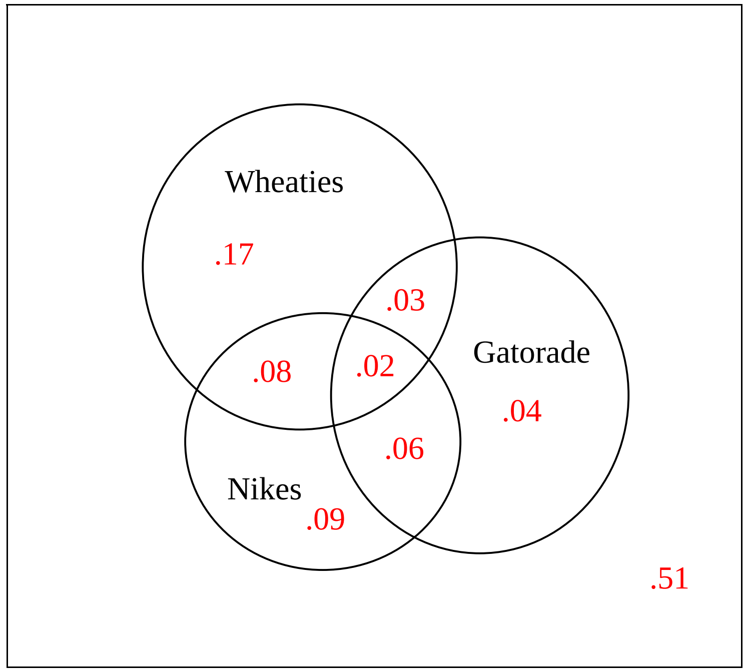
a) $P(G \cup W \cup N) = .49$

b) $P(G \cup W) = .40$

c) $P(N | G \cup W) = .16/.40 = .40$

d) $P(G \cap N^c) = .07$

e) $P(G | N^c) = .07/.75 = .0933$



Question 4

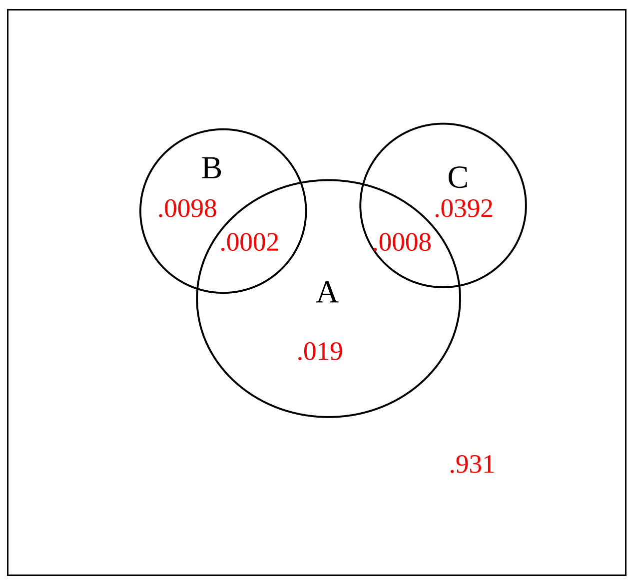
I call this my "Mickey Mouse" question.

Look at the Venn Diagram.

$P(A \cap B) = P(A)P(B|A) = P(A)P(B) = .0002$

$P(A \cap C) = P(A)P(C|A) = P(A)P(C) = .0008$

Ans: .069



Chapter 5

Question 1

a) $\mu = 3$ and $\sigma = \sqrt{2.6}$

b) Expected wealth is you take the offer is $250,000; expected wealth if you reject it is $300,000. If you care only about expected wealth, you reject it.

c) Expected utility if you take the offer is $E(U) = 10*2.5 - (2.5)^2 = 18.75$.

If you reject it and take the auction, your expected utility is

$$E(U) = 10E(x) - E(x^2) = 10\mu - (\mu^2 + \sigma^2)$$
$$E(U) = 10*3 - (9 + 2.6) = 18.4$$

Therefore you are better off accepting the buyout offer.

d) This person seems to be risk averse. They turn down an alternative with a bigger expected payoff in favor of one that has a smaller payoff but no risk.

Question 2

a) $\mu = 1, \sigma^2 = 1$

$y = a + bx = 100000 + 7500x$

b) $\mu_y = a + b\mu_x = 107500$

$\sigma_y^2 = b^2\sigma_x^2 = 7500^2 = 56{,}250{,}000$

Question 3

Ans: .3105 - .1628 = .1477

Question 4

$$f(0) = \binom{2}{0}(.1)^0(.9)^2 = .81$$

a) $$f(1) = \binom{2}{1}(.1)^1(.9)^1 = .18$$

$$f(2) = \binom{2}{2}(.1)^2(.9)^0 = .01$$

$$f(0)=\frac{\binom{2}{0}\binom{18}{2}}{\binom{20}{2}}=\frac{153}{190}=.8053$$

b) $$f(1)=\frac{\binom{2}{1}\binom{18}{1}}{\binom{20}{2}}=\frac{36}{190}=.1894$$

$$f(2)=\frac{\binom{2}{2}\binom{18}{0}}{\binom{20}{2}}=\frac{1}{190}=.0053$$

In part b the draws are not even approximately independent; getting a defective on the first draw lowers the chance of getting a second, while getting a non-defective on the first draw also lowers the chance of getting a second non-defective.

Question 5

$$f(0)+f(1)+f(2)$$

$$e^{-7.4}+\frac{7.4*e^{-7.4}}{1}+\frac{(7.4)^2 e^{-7.4}}{2!}$$

$$.0006+.0045+.0167=.0218$$

Question 6

a) $$f(3)=\frac{\binom{5}{3}\binom{29}{2}}{\binom{34}{5}}=\frac{4060}{278256}=.0145909$$

2) $$f(4)=\frac{\binom{5}{4}\binom{29}{1}}{\binom{34}{5}}=\frac{145}{278256}=.000521103$$

3) $$f(5)=\frac{\binom{5}{5}\binom{29}{0}}{\binom{34}{5}}=\frac{1}{278256}$$

4) The calculations below show the expected value of the ticket is about \$0.484, so that the average loss on each ticket is \$0.516.

x	f(x)	xf(x)
0	0.984884	0.000000
5	0.014591	0.072955
100	0.000521	0.052110
100,000	0.000004	0.359400
		0.484460

Question 7

Using the binomial, the chance of no blacks is $f(0)=\binom{50}{0}(.01)^0(.99)^{50}=.6050$,

and the chance of at least one is $1-f(0)=1-.6050=.3950$.

Using the poisson approximation, with $\mu=np=.50$, the probability of no blacks is $f(0)=\frac{(.5)^0e^{-.5}}{0!}=.6065$, and the probability of at least one is $1-f(0)=1-.6065=.3935$.

Question 8

a) First you need to show that $\mu_x=19, \mu_y=7$. Then we know that

$$\sigma_{xy}=\sum_x\sum_y(x-\mu_x)(y-\mu_y)f(x,y)$$

$$=(10-19)(6-7)(.3)+(20-19)(6-7)(.1)+(30-19)(6-7)(.1)+$$

$$(10-19)(8-7)(.1)+(20-19)(8-7)(.2)+(30-19)(8-7)(.2)=3.0$$

While $\rho_{xy}=\frac{\sigma_{xy}}{\sigma_x\sigma_y}=\frac{3.0}{\sqrt{1}\sqrt{69}}=.361158$.

2) $P(Y=6|X=10)=\frac{P(X=10\cap Y=6)}{P(X=10)}=\frac{.3}{.4}=.75$

3)

$$Z=.6X+.4Y$$

$$E(Z)=.6E(X)+.4E(Y)=.6(19)+.4(7)=14.2$$

$$\sigma_z^2=(.6)^2(69)+(.4)^2(1)+2(.6)(.4)(3)=26.44$$

$$\sigma_z=\sqrt{26.44}=5.142$$

Question 9

The number of points scored is given by: Points $=3F+7T$. Therefore the expected number of points is given by: $E(\text{Points})=3E(F)+7E(T)=3(2.1)+7(1.7)=18.2$.

To get the standard deviation we must first get the covariance. We know

$.30 = \rho_{TF} = \frac{\sigma_{TF}}{\sigma_T \sigma_F} = \frac{\sigma_{TF}}{1(1.2)} \Rightarrow \sigma_{TF} = .36$. Therefore, we can get the variance of points:

$$\sigma_P^2 = (3)^2(1.2)^2 + (7)^2(1)^2 + 2(3)(7)(.36) = 77.08$$

$$\sigma_p = \sqrt{77.08} = 8.78$$

Chapter 6

Question 1

4) $\int_0^1 2x\,dx = \Big|_0^1 x^2 = 1 - 0 = 1$

5) $\int_0^{.5} 2x\,dx = \Big|_0^{.5} x^2 = .25 - 0 = .25$

6) $\int_{.8}^1 2x\,dx = \Big|_{.8}^1 x^2 = 1 - .64 = .36$

7) $\int_{.5}^{.8} 2x\,dx = \Big|_{.5}^{.8} x^2 = .64 - .25 = .39$

Question 2

a) True by symmetry.
b) False; equally likely by symmetry.
c) True.
d) False; $P\left(z < \frac{65-45}{10} = 2.0\right) = .9772 \neq .955$.

Question 3

$$P(X > 150{,}000) = P\left(z > \frac{150{,}000 - 120{,}000}{10{,}000} = 3\right) = .0014$$

Question 4

a) $P(X < 500) = P\left(z < \frac{500-500}{90} = 0\right) = .5000$

b) $P(320 < X < 680) = P\left(-2 = \frac{320-500}{90} < z < \frac{680-500}{90} = 2\right) = .9544$

c) $\frac{X^* - 500}{90} = -z_{.15} = -1.04$
$X^* = 500 - 90(1.04) = 406.4$

Question 5

a) Poisson – $f(1) = \frac{(2)^1 e^{-2}}{1!} = .2707.$

b) Exponential $P(1 < x < 2) = \left(1 - e^{-\frac{2}{.5}}\right) - \left(1 - e^{-\frac{1}{.5}}\right) = e^{-2} - e^{-4} = .1353 - .0183 = .1170$

Chapter 7

Question 1

a)

$$P\left(\sum x_i > 97.5(12) = 1170\right) = ?$$

$$P\left(\frac{\sum x_i}{600} = \overline{X} > \frac{1170}{600} = 1.95\right) = ?$$

$$E\left(\overline{X}\right) = \mu = 2.0, \quad \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{1}{\sqrt{600}} = .0408$$

$$P\left(\overline{X} > 1.95\right) = P\left(z > \frac{1.95 - 2.00}{.0408} = -1.22\right) = .8888$$

b) The central limit theorem is relevant because book widths are not assumed to be normally distributed. However, we could use the normal distribution to make probability statements about the sample mean because 600 is a large sample size, and sample means from large samples are approximately normally distributed for (almost) any parent distribution.

Question 2

a) First we need to calculate $\mu = 10, \sigma = 7.642$.

$$P\left(\sum x_i > X^*\right) = .01$$

$$P\left(\overline{X} = \frac{\sum x_i}{15000} > \frac{X^*}{15000}\right) = .01$$

$$E\left(\overline{X}\right) = 10, \; \sigma_{\overline{X}} = \frac{7.642}{\sqrt{15000}} = .0624$$

$$\frac{\frac{X^*}{15000} - 10}{.0624} = z_{.01} = 2.33$$

$$X^* = 15000\left(10 + (.0624)(2.33)\right) = 152{,}181$$

b)

$$E\left(\overline{p}\right) = p = .7, \; \sigma_{\overline{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(.7)(.3)}{15000}} = .00374$$

$$P\left(\overline{p} > .69\right) = P\left(z > \frac{.69 - .70}{.00374} = -2.67\right) = .9962$$

Question 3

1) $$P(.67 < \bar{p} < .72) = P\left(\frac{.67-.70}{\sqrt{\frac{(.7)(.3)}{1000}}} < z < \frac{.72-.70}{\sqrt{\frac{(.7)(.3)}{1000}}}\right)$$

$$P(-2.07 < z < 1.38) = .4808 + .4162 = .897$$

2) $$-z_{.10} = -1.28 = \frac{.64-.70}{\sqrt{\frac{(.7)(.3)}{n}}}$$

$$n = 95.6 \Rightarrow n = 96$$

Question 4

a -- Agree; these are just two ways of describing the same event.

B – Disagree; while the standard deviation of a proportion of successes is $\sqrt{\frac{p(1-p)}{n}}$ and therefore goes to zero as n increases to infinity, the standard deviation of the *number* of successes is $\sqrt{np(1-p)}$ which gets indefinitely large as n increases to infinity.

C – Disagree; the Central limit theorem says we can use the normal distribution to approximate the distribution of sample means constructed from large samples. It does not say we can use the normal distribution to make probability statements about single observations.

D – Agree.

E – Disagree. Only if the parent distribution is normal will the sample means always be exactly normally distributed.

Question 5

a – Both estimators seem unbiased. If there is a bias it is very small. This is because the histograms look as if they are centered on .5 (the true population mean), and the average values given by the describe command are .50511 and .51595, which are fairly close to .500. It also looks like the sample mean is the more efficient estimator. In the pictures, the spread is obviously less for the sample means, and this is confirmed by the standard deviations .11502 versus .16844.

b) $$\text{MSE} = (\text{bias})^2 + \text{variance}$$

$$\text{MSE}_{\bar{X}} = (.50511\text{-}.5)^2 + (.11502)^2 = .00003 + .01323 = .01326$$

$$\text{MSE}_{\text{MED}} = (.51595 - .5)^2 + (.16844)^2 = .00025 + .02837 = .02862$$

$$\text{RELATIVE EFFICIENCY} = .02862/.01326 = 2.15$$

Question 6

$$
\begin{aligned}
E\left(\hat{\theta}\right) &= k\left\{\left(\sum_{i=1}^{n} E\left(x_i^2\right)\right) - nE\left(\bar{x}^2\right)\right\} \\
&= k\left\{\left(n\left(\sigma^2 + \mu^2\right)\right) - n\left(\mu^2 + \frac{(N-n)\sigma^2}{(N-1)n}\right)\right\} \\
&= k\left\{n - \frac{(N-n)}{(N-1)}\right\}\sigma^2 \\
&= k\left\{\frac{N(n-1)}{(N-1)}\right\}\sigma^2 \Rightarrow k = \frac{(N-1)}{N(n-1)}
\end{aligned}
$$

Chapter 8

Question 1

Since $\bar{X} + t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} = 1.27 + t_{\frac{\alpha}{2}} \frac{1.126}{\sqrt{10}} = 2.2743$

It must be that $t_{\frac{\alpha}{2}} = \dfrac{2.2743 - 1.27}{\frac{1.126}{\sqrt{10}}} = 2.82$. Using a t-table with 9 d.f. we see that

$\frac{\alpha}{2} = .01 \Rightarrow (1-\alpha) = .98$

Question 2

No. This implies the population mean is a random variable, and is jumping around, and is between 3.8 and 12.2 95% of the time. The population mean is a fixed number, and it is either between 3.8 and 12.2 or it isn't. The correct interpretation is that we used a technique which, if used over and over again, would bracket the true population mean for 95% of all samples drawn. We have no way to know if this is one of the samples that works, or one that doesn't.

Question 3

a) False.
b) False and silly. There is only one population mean.
c) True.
d) False and silly. Families don't have means.
e) False. Exactly 100%, by construction, will bracket the SAMPLE mean.
f) True

Chapter 9

Question 1

The boundary of the acceptance rejection region is given by the formula

$$z_{.01} = \frac{\overline{X}^* - 30}{1.5/\sqrt{32}} \Rightarrow \overline{X}^* = 30 + 2.33 \times 1.5/\sqrt{32} = 30.62$$

Since the observed sample mean is 30.8, we would reject the null hypothesis.
The power is the probability of correctly rejecting the null when the null is false. In this case

$$P\left(\overline{X} > 30.62 \mid \mu = 31\right) = P\left(z > \frac{30.62 - 31}{1.5/\sqrt{32}}\right) = P(z > -1.43) = .9236$$

Question 2

Alpha is the chance of rejecting when the null is true. If the null is true, then there are two red balls in the urn, out of seven. The probability of rejecting is the probability (given by the hypergeometric distribution of drawing both red balls.)

$$f(2) = \frac{\binom{2}{2}\binom{5}{0}}{\binom{7}{2}} = \frac{1}{21}$$

The chance of a type 2 error is the chance that you *don't* get two red balls when there are four in the urn. This is also hypergeometric.

$$1 - f(2) = 1 - \frac{\binom{4}{2}\binom{3}{0}}{\binom{7}{2}} = 1 - \frac{6}{21} = \frac{15}{21}$$

Question 3

a) $H_0 : p \le \frac{1}{6}$

$H_A : p > \frac{1}{6}$

b) This is binomial. $\sum_{x=6}^{15} \binom{15}{x} \left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{15-x}$.

$$E\left(\bar{p}\right) = p = .1666$$

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.16666 \times .83333}{15}} = .0962$$

c) We approximate this binomial with a normal.

$$\bar{p} = \frac{6}{15} = .40$$

$$z = \frac{.40 - .166}{.0962} = 2.43$$

The probability that z is bigger than 2.43 is .0075
d) Again, binomial

$$\sum_{x=0}^{5} \binom{15}{x} (.3)^x (.7)^{15-x}$$

e) cheating, honest; honest, cheating

Question 4

You need a picture, and the definition of the p-value. The test performed was a less than test, so the area to the left of 10.233 is .89. Therefore the right tail area is .11. For a two-tailed p-value, the area we want is the sum of the two (symmetric) tails, so the answer is .22.

Question 5

This data is real, and is the major study supporting the American Heart Association's "know your number" campaign. The calculations show that the evidence supporting the campaign is not as convincing as you might have been led to believe.
a) If the treatment were ineffective, half the heart attacks should come in the treatment group, and half in the control group (because they are the same size). If the treatment lowers heart attack risk, we'd expect fewer than half the heart attacks in the treatment group.

$$H_0 : p \geq .50$$
$$H_A : p < .50$$

where p is the population proportion of heart attacks in the treatment group.

b) In our sample, $\bar{p} = \frac{160}{356} = .4494$. The p-value for the test is

$$P\left(z < \frac{.4494 - .50}{\sqrt{\frac{(.5)(.5)}{356}}} = -1.91 \right) = .0281.$$

This would be enough to reject the null hypothesis at using a 5% significance level.
c) However, if you had done a two-tailed test, the p-value would be twice as big, or .0562, in which case the result is *not* statistically significant at the 5% level. The American Heart Association intended to do a two-tailed test, until it looked at the data and discovered it gave the "wrong" answer, at which point they went to a one-tailed test. Switching the type of the test based on the sample data isn't kosher. An extremely thought-provoking article in the *Atlantic Monthly* in September 1989 drew attention to this and other weaknesses in the case against

cholesterol. You can explore this on the web if you wish by visiting http://heskco.com/karan/myth.htm

Question 6.

a) $H_0: \mu \geq 32$
$H_A: \mu < 32$

b) $H_0: p_{ny} = p_{nj}$
$H_A: p_{ny} \neq p_{nj}$

c) $H_0: \mu = 2$
$H_A: \mu \neq 2$

d) $H_0: p \leq .10$
$H_A: p > .10$

Question 7.

Despite the apparently small p-value, the statistical evidence that asbestos in the drinking water causes cancer is weak. The problem is that even with a p-value of .001, there is still one chance in a thousand of getting a rejection by chance. (This is what it means to have a p-value of one in a thousand.) If you perform 200 different tests, those one-in-a-thousand chances start to add up. If the tests were all statistically independent of one another (which is unlikely, but allows me to illustrate how the small chances add up) the chance of getting at least one success in 200 trials could be computed using the binomial distribution, with n=200 and p=.001. In that particular case, the chance of one or more results “statistically significant at the .001 level” is 18.14%. Since the likelihood of the tests being independent is not high, we can’t actually compute the correct p-value, but it is likely to be much nearer 0.18 than 0.001. Then there is an additional problem: namely, that smoking was not held constant. Since smoking is generally regarded to be an important cause of lung cancer, and smoking can be correlated with important included variables, estimates that fail to hold smoking constant are suspect. It may simply be that people who live in neighborhoods with asbestos in the water are also from socio-demographic groups more inclined to smoke.

Then there is the important distinction between being “statistically significant” and “practically important.” Statistically significant means that a difference can be perceived using statistical methods: it does not mean the difference is large or especially important. In this case, the point estimate of the size of the effect suggests a hundredfold increase in asbestos increases cancer risk by only 1.05. Practically speaking, this is not a strong effect, even if one ignores the objections in the first paragraph, which suggest there may be no effect at all.

Chapter 10

Question 1

This is a paired sample, one-sided, difference of two means test. The correct answer (which I got from Minitab) is:

Paired T-Test and Confidence Interval

```
Paired T for C2 - C1

                  N       Mean     StDev   SE Mean
C2                5     114.00     13.82      6.18
C1                5     109.00      8.37      3.74
Difference        5       5.00     10.12      4.53

95% CI for mean difference: (-7.57, 17.57)
T-Test of mean difference = 0 (vs > 0): T-Value = 1.10  P-Value = 0.166
```

The p-value of .166 > .10, so you accept the null that college does not improve IQ.
Done by hand, you get:

	With College	Without College	Difference
Smith	118	112	6
Jones	105	98	7
Brown	135	120	15
Hall	99	111	-12
Davis	113	104	9

$$\bar{d} = \frac{\sum_i d_i}{n} = \frac{25}{5} = 5$$

$$s_d = \sqrt{\frac{\sum_i \left(d_i - \bar{d}\right)^2}{n-1}} = \sqrt{\frac{410}{4}} = 10.12$$

$$t_{obs} = \frac{\bar{d} - 0}{s_d / \sqrt{n}} = \frac{5-0}{10.12/\sqrt{5}} = 1.10 < t_\alpha = 1.533$$

Question 2

This is a test of the difference of two means, using independent samples.
You get:

Sports Illustrated	*Scientific American*
$\bar{x} = \frac{\sum x_i}{n} = 6.7383$	$\bar{x} = \frac{\sum x_i}{n} = 10.968$
$s = \sqrt{\frac{(x_i - \bar{x})^2}{n-1}} = 1.6356$	$s = \sqrt{\frac{(x_i - \bar{x})^2}{n-1}} = 2.6467$
$n = 6$	$n = 6$

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{(1.6356)^2}{6} + \frac{(2.6467)^2}{6}} = 1.27018$$

$$t_{obs} = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{10.968 - 6.783}{1.27018} = 3.2999$$

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2} = \frac{2.603}{.3124} = 8.33$$

$$3.2999 > t_{.05} = 1.860$$

Therefore, you reject the null hypothesis.

Chapter 14, Part I

Question 1

Here is the minitab output, which can be used to check your hand calculations.

Regression Analysis

```
The regression equation is
Injury% = - 8.46 + 0.267 temp

Predictor        Coef       StDev          T        P
Constant       -8.460       2.333      -3.63    0.011
temp          0.26719     0.03897       6.86    0.000

S = 1.027       R-Sq = 88.7%     R-Sq(adj) = 86.8%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         1      49.607      49.607     47.01    0.000
Residual Error     6       6.332       1.055
Total              7      55.939


Predicted Values

     Fit  StDev Fit         99.0% CI             99.0% PI
   8.907      0.429   (   7.315,  10.499)  (   4.779,  13.035)
```

a) `Injury% = - 8.46 + 0.267 temp`
b) `1.055`
c) `R-Sq = 88.7%`
d) `(   4.779,  13.035)`

Question 2

First, note that $\pm 1.96 s_{b_1} = \pm .27$, which means that $s_{b_1} = \frac{.27}{1.96} = .1378$. Therefore the z-value (it is a z value because of the huge sample size) for this null hypothesis is $z = \frac{.38 - 0}{.1378} = 2.76$ and the p-value on a "greater than" test is .0029.

Question 3

a) The error variance seems to be increasing with x, because the spread of observations is broader at big values of x than at small values of x.
b) This is harder to see. The errors are not independent. Positive residuals are most likely to be followed by other positive residuals; negative residuals by negative residuals. The residuals meander around zero.

c) The concern is the influential observation. The diagnostic statistic is the leverage, computed with formula 14.33 in the text. Another statistic, formula 15.25, is called Cook's distance, but it doesn't get introduced until chapter 15. One experiment you'd want to try (after confirming the extreme observation was not simply an error) is to fit the regression with and without the observation to see how much difference it makes.

Chapter 15

Question 1

a) .178
b) .064
c) .87
d) .062
e) For (a), … you'd get a coefficient as big or bigger than 635.4, when the null hypothesis is true.; for (b) . . you'd get a coefficient as far or further from zero as the 129.1 you observed, when the null hypothesis is true; for (c) … you'd get a coefficient as big or bigger than –5.85, when the null hypothesis is true.; for (d) … you'd get a coefficient as small or smaller than –9.37 when the null hypothesis is true.

Question 2

a) 9.09, 31.95, 130.4, 12.9
b) The observed t is 1.78, the critical t is 1.325. This assumes there is actually a constant in the model, even though it isn't mentioned on the printout, so degrees of freedom = 20. Thus you reject the null of no difference in the sexes.
c) 180/39 = 4.61 years. This is only approximate because both 180 and 39 are sample estimates, and not the true underlying population parameters.

Question 4

One of the most important things to remember about hypothesis tests is that *Accepting the null hypothesis does not mean you have proven it is true. It only means you have not found compelling evidence to refute it*. The political scientists who wrote "inflation has no impact on voting behavior" apparently forgot this. A correct interpretation of the result is that the statistical evidence suggests inflation increases the Republican vote by about seven percent (an amount large enough to have important electoral consequences) but that there is so much uncertainty surrounding the estimate that the effect could be substantially larger or smaller, and could even be zero.

Chapter 16

Question 5

a) False. The size of the coefficients depends on the units used to measure the variables. For example, consider the coefficient of DRLATE, which is .09. Had DRLATE been measured in *hours* instead of *minutes*, its coefficient would have been 60x.09 = 5.4. Another way to see that SHORT isn't a key factor is by noting the average wait is 32 minutes and the standard deviation of the wait is 15 minutes. SHORT accounts for only 2.61 minutes of waiting, which is small compared to both the mean and standard deviation of waiting times.

b) False. 5.4 minutes.

c) i) False. It should be $\mu = 32 \pm 1.96\frac{15}{\sqrt{200}}$, NOT $\mu = 32 \pm 1.96(15) \approx 32 \pm 30$

ii) False. The 95% C.I. is roughly $\begin{array}{c} -.24 \pm (1.96)(.05) \cong .24 \pm .10 \\ -.14 \text{ to } -.34 \end{array}$

d) False. The coefficient of PALATE is negative. Arriving late means a *shorter* wait.

e) False. True for R^2 but not adjusted R^2.

f) Let $x_4 = \begin{cases} 1 \text{ if Betty} \\ 0 \text{ otherwise} \end{cases}$ $x_5 = \begin{cases} 1 \text{ if Barney} \\ 0 \text{ otherwise} \end{cases}$ $x_6 = \begin{cases} 1 \text{ if Wilma} \\ 0 \text{ otherwise} \end{cases}$

Patient	X4	X5	X6
A	1	0	0
B	0	1	0
C	0	0	1
D	0	0	0

$F_{obs} = \frac{(12,537 - 6,742)/3}{6,742/193} = \frac{1931.7}{34.93} = 55.30$ Since this is bigger than the critical F value using $\alpha = .01$ (namely, $F_{.01,3,193} = 4.00$) the result is clearly statistically significant.