8.6 (3 points)
a)   Using p-hat, the sample proportion p-hat = 524/1711 = .3167
with SE = .01125. Using the Wilson estimate, p-tilde = 526/1715
= .3172, with SE = .01125. These are essentially the same.
b)   The 99% interval is .2878 to .3457
c)   No, we do not know what fraction of the cyclist who are not
involved in accidents had alcohol in their system.

8.20 (2 points)
a)   Because 19 preferred instant coffee, we assume the other 31
preferred freshly brewed. The sample proportion is .62 with a
SE of 0.0707 $(= \sqrt{\dfrac{0.5 \times 0.5}{50}})$. The Z statistic is 1.697 and the
one-sided p-value is .0446. This is significant at 5%. We
conclude that this is strong evidence that freshly brewed is
preferred over instant.
b)   The 90% interval is .5071 to .733. This shows that there
is a chance that the "majority" could be as slim as a bare 51%
of the population.

8.36 (2 points)
a)   The two estimates are .1451 (female) and .3390 (male). The
SE of the difference is .0281 and the 95% interval is
from .13887 to .24903 (male-female).
b)   The term for female is much larger than for male. Although
the estimate for female is closer to 0, which would make its SE
smaller, the n for female is so much smaller than for male that
the resulting term is more than four times as large.

8.46 (4 points)
a)   The overall proportion who support the proposal is .454
b)   The SE is .0295
c) H_0: p_1 = p_2 vs. H_a: $p_1 \neq p_2$
d)   Z = 3.4747 and the two-sided p-value is .0005. This is
quite strong evidence that the support differs in the two
counties.

9.28 (3 points)
a)   Column percents because "source" is the explanatory table.
These are in the following table. For Cases, Private and Pet
store are similar, but Other is rather smaller.

	Private	Pet Store	Other
Cases	36.2	40.0	27.2
Controls	63.8	60.0	72.8

Total	100	100	100

b)   The following table gives expected values.

	Private	Pet Store	Other	Total
Cases	111.915	13.0514	91.0332	216
Controls	231.085	26.9486	187.967	446
Total	343	40	279	662

$\chi^2 = 6.61$, df = 2, p-value = 0.037 which supports a difference between the cases and controls.

9.34 (4 points)

a)

	Regular	World Series	Total
Hit	2584	35	2619
No Hit	7280	63	7343
Total	9864	98	9962

b)

	Regular	World Series
Hit	26.2	35.7
No hit	73.8	64.3
Total	100	100

c)

H_0: "Yes" rate is the same for both vs. H_a: the two rates are not the same. Expected values are given below.

	Regular	World Series	Total
Hit	2593.24	25.7641	2619
No hit	7270.76	72.2359	7343
Total	9864	98	9962

Following table gives terms in χ^2 statistic. The value of the statistic is 4.536 with df = 1. The p-value is between .05 and .025. This is fairly strong evidence that the two rates are different.

	Regular	World Series	Total
Hit	.03289	3.31088	3.34377
No hit	.01173	1.18088	1.19261
Total	.04463	4.49176	4.53638

9.35 (3 points)

The column percents are below. Almost 70% of next year's winners are winners this year, while only 30% of next year's winners were losers next year.

	Winner	Loser
Winner	69.7	29.7
Loser	30.3	70.3
Total	100.0	100.0

H_0: "Yes" rate is the same for both vs. H_a: The two rates are not the same. Expected values are given below.

	Winner	Loser	Total
Winner	61	59	120
Loser	61	59	120
Total	122	118	240

The following table gives terms in χ^2 statistic. The value of the statistic is 38.41 with df = 1. This is extremely significant. This is extremely strong (overwhelming) evidence that the two rates are different.

	Winner	Loser	Total
Winner	9.44262	9.76271	19.2053
Loser	9.44262	9.76271	19.2053
Total	18.8852	19.5254	38.4107

9.36 (2 points)
The pooled estimate is ½ (because we classified based on above or below the median). The SE for the difference is .64559 and Z = 6.1976. Z^2 = 38.41

9.46 (3 points)
The column percents tell what percent of those with loans entered each field, as well as the percent of those without loans who went into the same fields. A slightly lower fraction of those with loans went into Management than did those without loans, while a slightly higher fraction of those went into Science than did those without loans. The other percents seem quite close.

	Loan	No loan	Total
Agriculture	8.7	7.0	7.7
Child Develop.	10.1	10.1	10.1
Engineering	26.6	27.6	27.2
Liberal Arts	24.2	24.9	24.6
Management	6.5	10.3	8.7
Science	8.4	5.8	6.9
Technology	15.5	14.3	14.8

The expected values are below.

	Loan	No loan	Total
Agriculture	28.504	38.496	67
Child Develop.	37.0127	499.9873	87
Engineering	99.9769	135.023	235
Liberal Arts	90.6173	122.383	213
Management	31.9075	43.0925	75
Science	25.526	34.474	60
Technology	54.4555	73.5445	128
Total	368	497	865

The following table gives terms in χ^2 statistic. The value of the statistic is 6.525 with df = 6. This is extremely insignificant.
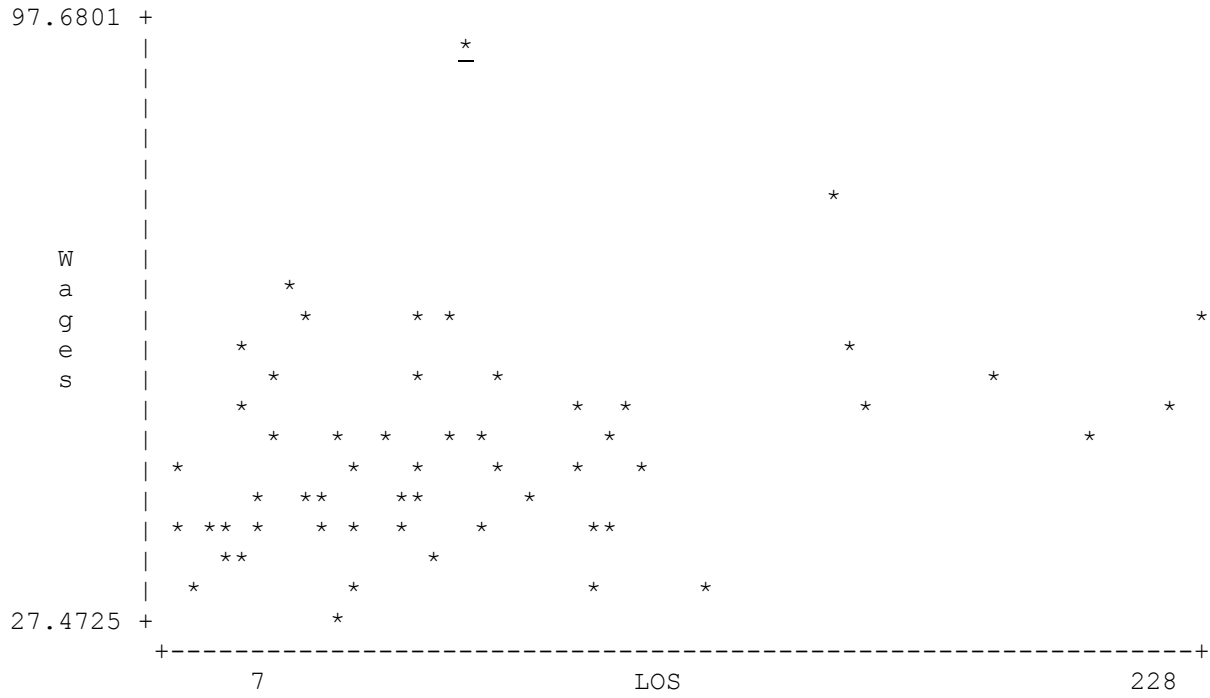
	Loan	No loan	Total
Agriculture	.42877	.31748	.74625
Child Develop.	4.4E-06	3.2E-06	7.6E-06
Engineering	.03909	.02894	.06803
Liberal Arts	.02887	.02137	.05024
Management	1.95969	1.45104	3.41073
Science	1.17388	.86919	2.04308
Technology	.1189	.08804	.20693
Total	3.7492	2.77607	6.52526

10.2 (3 points)
$y = \beta_0 + \beta_0 x + \varepsilon$
a) β_0 is the fixed expenses term.
b) β_1 allows for change when x changes. We expect it to be positive because costs will increase with the number of students.
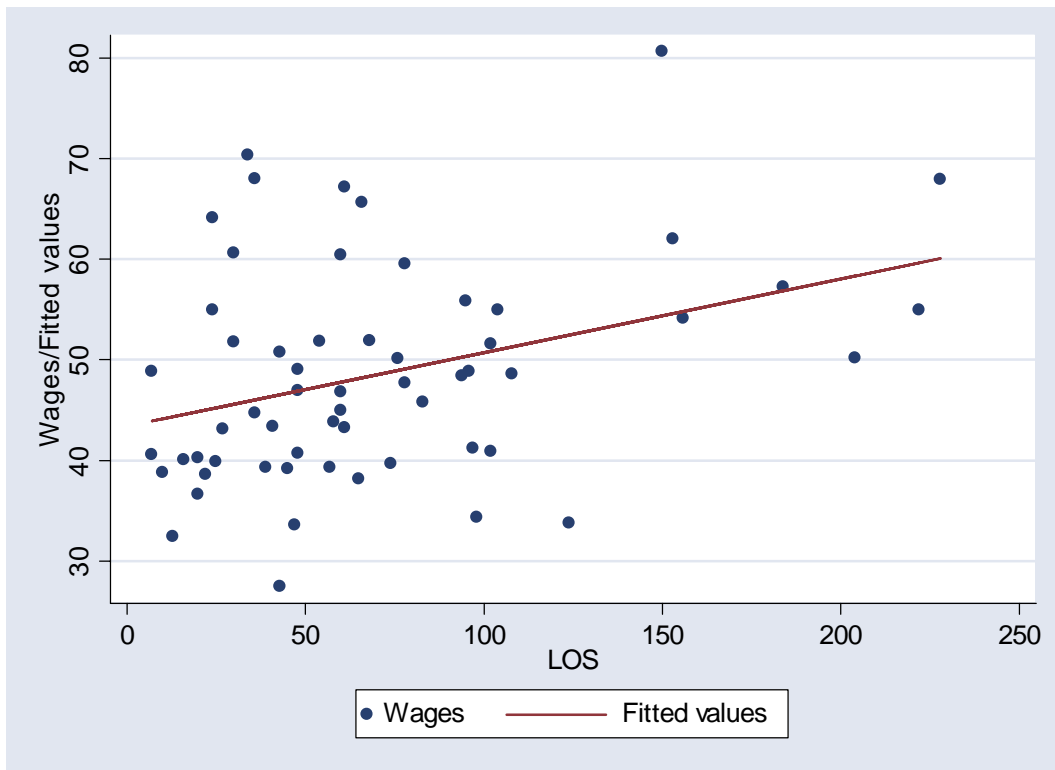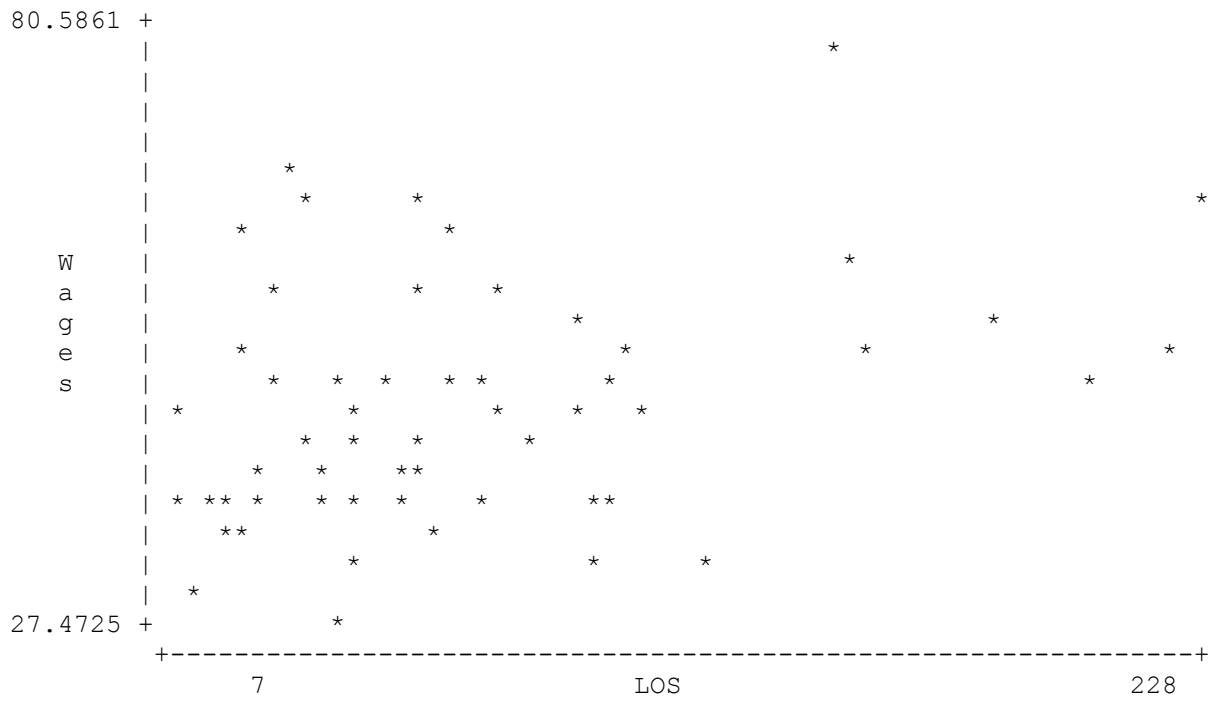c) ε

10.9 (4 points)
. plot wages los

```
  97.6801 +
          |
          |                              *
          |                             ━
          |
          |
          |
          |                                                  *
          |
        W |
        a |                  *
        g |             *                                            *
        e |       *         *    *  *
        s |                                              *
          |          *     *       *      *                      *
          |             *      *  *      * *           *                 *
          |      *            *       *        *             *        *
          |   *         *   *   *      *  *      *    *      *
          |         *    **       **         *
          |   * **  *      *  *    *     *        **
          |     **               *
          |   *              *              *          *
  27.4725 +           *
          +----------------------------------------------------------+
                7                        LOS                      228
```

. summarize wages

Variable	Obs	Mean	Std. Dev.	Min	Max
wages	60	49.3651	12.46447	27.4725	97.6801

. drop if wages > 97
(1 observation deleted)

. plot wages los

```
80.5861 +                                                        *
         |
         |
         |
         |                                                                    *
         |        *                                                *
         |         *         *
W        |       *         *                  *
a        |            *                                  *
g        |                       *       *                      *
e        |         *                                  *            *
s        |             *    *    *   *  *                  *           *
         |     *               *                *      *
         |              *   *    *       *                  *
         |   *          *  *     *  *  *      *        * *
         |       * *                     *
         |            *              *
27.4725 +    *          *
         +-------------------------------------------------------------------+
              7                            LOS                         228
```

b)
y = 43.4 + 0.0733x.   The t statistic is 2.85 with df = 57.   The
p-value is .006.   The slope is significantly different from 0.
c)   The slope tells us that wages increase by .07325 for each
increase of 1 in LOS.   Without the units, we cannot convert this
into a statement of dollars per week.
d)   99% interval is .004748 to .14175

10.20 (5 points)
a)   There are no outliers, although some of the points for low
HR are further from the line than the other points.
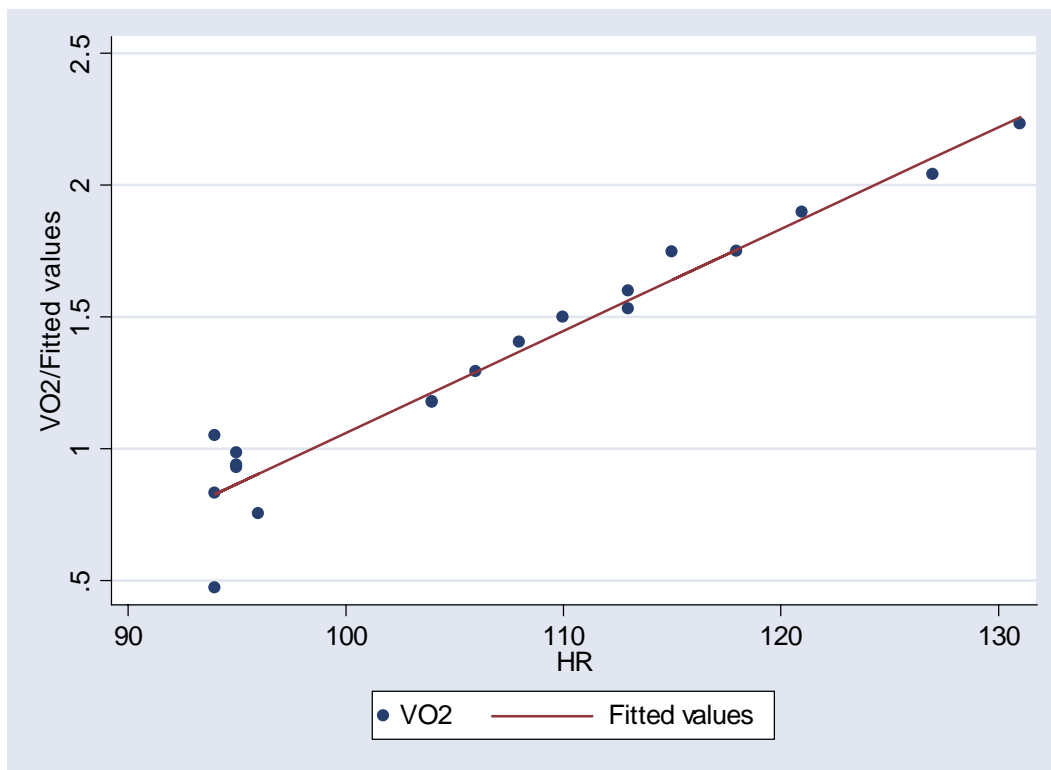
b) y = -2.80 + 0.0387x
c)   The t statistic is 16.10 with df = 17.   The p-value is
essentially 0.   This means that we are quite sure that there is
a linear relationship between V02 and HR.
d)   For HR = 96, the predicted value is .9062 and the 95%
prediction interval is .6396 to 1.1729.   For HR = 115, the
predicted value is 1.6406 and the 95% prediction interval is
1.3768 to 1.9045.
e)   It depends on how accurately they need to know V02.   The
regression equation predicts the mean value.   The prediction
interval show that there is considerable variation among
individuals.

10.32 (5 points)
a) Minitab output below

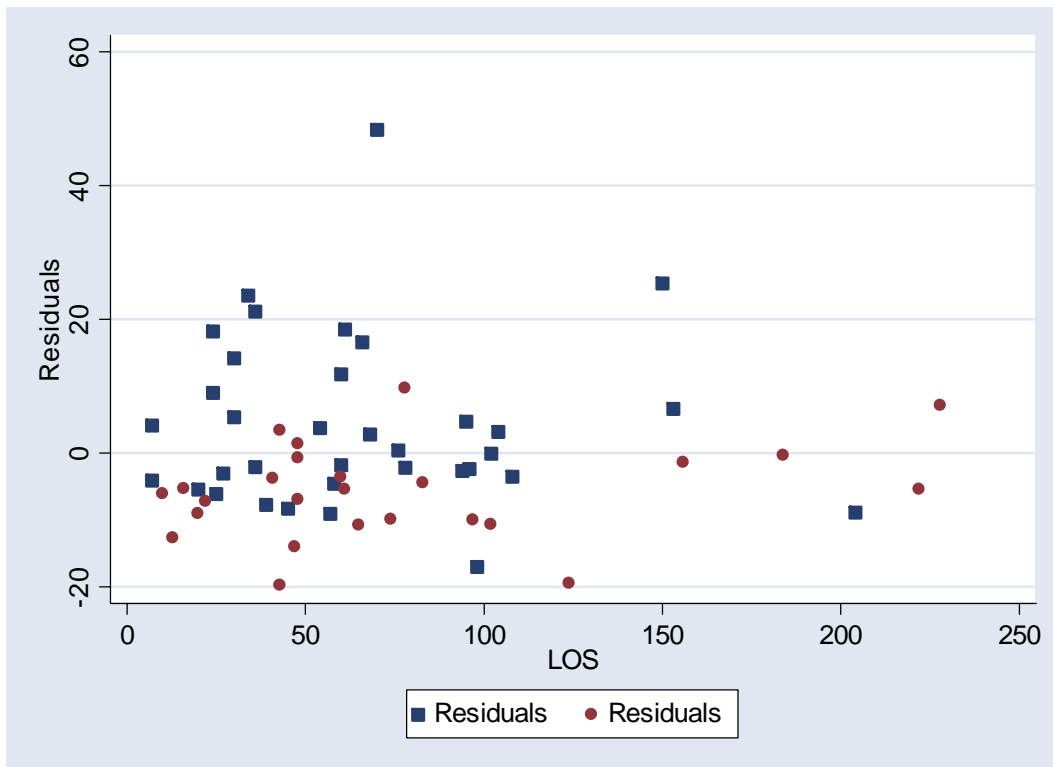| Source	DF	SS	MS	F	P
Regression	1	3.7619	307619	259.27	0.00
Residual Error	17	0.2467	0.0145		
Total	18	4.0085			

b)   This tests if the slope is 0, ie whether VO2 is unrelated
     to HR.
c)   If H_0 is true, then F has an $F(1,17)$ distribution. The
     p-value is <.0001
d)   We found t=16.10 and t^2 = 259.21
e) R^2 = SSM/SST = 3.7619/4.0085 = 93.8%

10.34 (2 points)
a) $t = \dfrac{-0.19\sqrt{711}}{\sqrt{1-(-0.19)^2}} = -5.160$

b)   df=711, p-value < 0.001. Conclude that $\rho \neq 0$.

10.40 (2 points)

Square is large and circle is small

Most small banks have negative residuals, while large banks have
mostly positive residuals. This means that, generally wages at

large banks are higher and small bank wages are smaller than we would predict from the regression.