Booklet of Code and Output
for
STAC32 Final Exam

December 12, 2015

List of Figures in this document by page:

List of Figures

```
degree,days
business,136
business,162
business,135
business,180
business,148
business,127
business,176
business,144
computer science,156
computer science,113
computer science,124
computer science,128
computer science,144
computer science,147
computer science,120
engineering,126
engineering,151
engineering,163
engineering,146
engineering,178
engineering,134
```

Figure 1: Time in days for students of different majors to find full-time employment

```
R> blueberry.1=aov(yield~variety,data=blueberry)
R> summary(blueberry.1)

            Df Sum Sq Mean Sq F value Pr(>F)
variety      3 0.0104 0.00348   0.085  0.968
Residuals   28 1.1449 0.04089
```

Figure 2: Analysis of variance of blueberry data

```
R> TukeyHSD(blueberry.1)

  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = yield ~ variety, data = blueberry)

$variety
                   diff        lwr       upr     p adj
Duke-Berkeley   -0.0425 -0.3185495 0.2335495 0.9745322
Jersey-Berkeley -0.0400 -0.3160495 0.2360495 0.9785924
Sierra-Berkeley -0.0125 -0.2885495 0.2635495 0.9993077
Jersey-Duke      0.0025 -0.2735495 0.2785495 0.9999944
Sierra-Duke      0.0300 -0.2460495 0.3060495 0.9907121
Sierra-Jersey    0.0275 -0.2485495 0.3035495 0.9928045
```

Figure 3: Tukey's method applied to blueberry data

```
R> boxplot(yield~variety,data=blueberry)
```

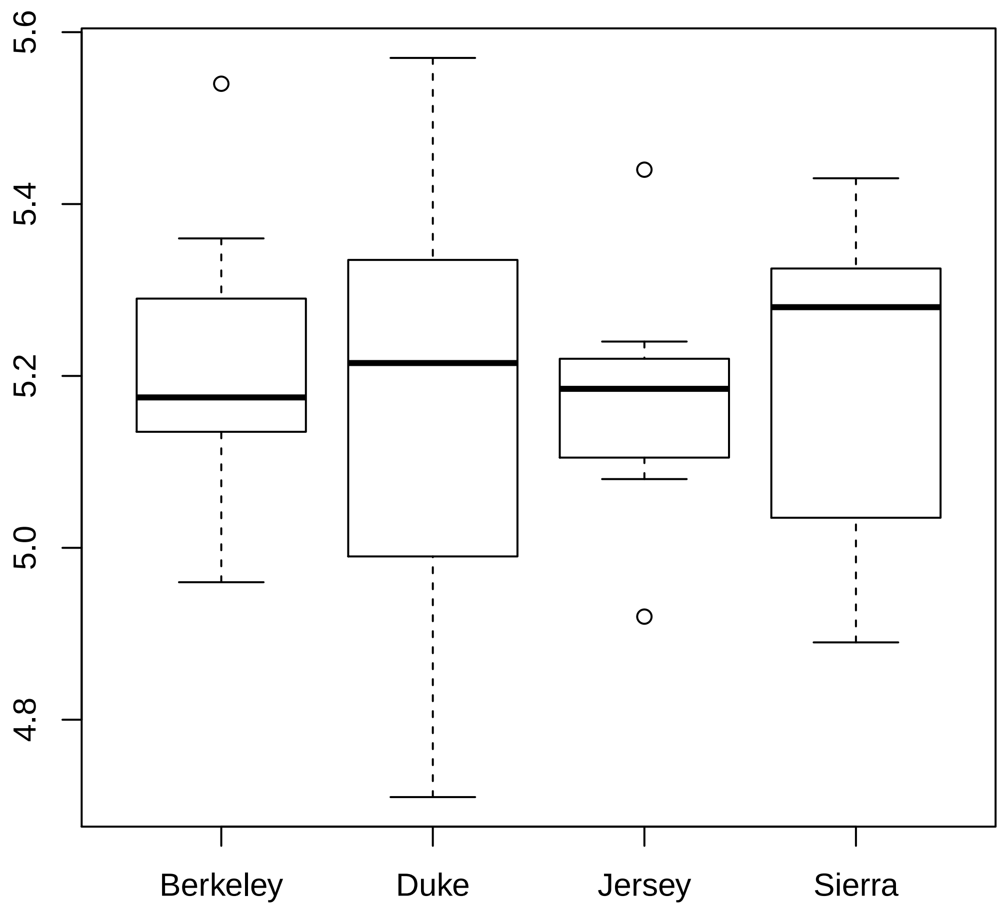

Figure 4: Boxplots of blueberry yields by variety

```
R> moissanite=read.table("moissanite.txt",header=T)
R> m.1=lm(VOLUME~PRESSURE,data=moissanite)
R> summary(m.1)

Call:
lm(formula = VOLUME ~ PRESSURE, data = moissanite)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.7765 -0.3549 -0.2123  0.2853  1.3851 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 98.614919   0.403723  244.26  < 2e-16 ***
PRESSURE    -0.255594   0.008646  -29.56 2.83e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6484 on 9 degrees of freedom
Multiple R-squared:  0.9898,	Adjusted R-squared:  0.9887 
F-statistic: 873.9 on 1 and 9 DF,  p-value: 2.832e-10
```

Figure 5: Linear regression for moissanite data

```
R> r=resid(m.1)
R> f=fitted(m.1)
R> plot(r~f)
R> lines(lowess(r~f))
```

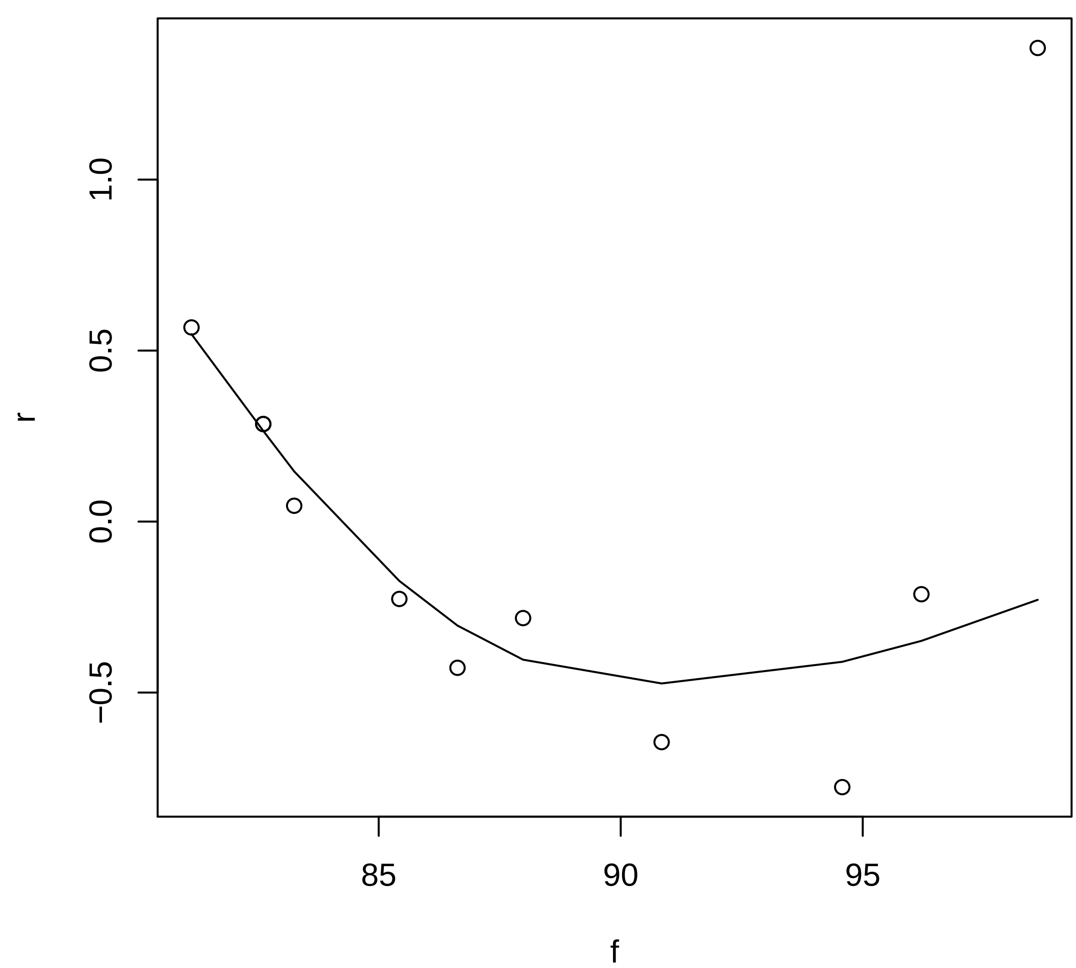

Figure 6: Residual plot for linear regression of moissanite data

```
R>   attach(moissanite)
R>   psq=PRESSURE^2
R>   m.2=lm(VOLUME~PRESSURE+psq)
R>   summary(m.2)
R>   detach(moissanite)

Call:
lm(formula = VOLUME ~ PRESSURE + psq)

Residuals:
     Min       1Q   Median       3Q      Max
-0.54312 -0.12355  0.00034  0.11405  0.49717

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 99.502832   0.268286  370.88  < 2e-16 ***
PRESSURE    -0.347272   0.018306  -18.97 6.17e-08 ***
psq          0.001311   0.000254    5.16 0.000864 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3305 on 8 degrees of freedom
Multiple R-squared:  0.9976,       Adjusted R-squared:  0.9971
F-statistic:  1694 on 2 and 8 DF,  p-value: 3.077e-11
```

Figure 7: A second regression for moissanite data

```
SAS> proc power;
SAS>   onesamplemeans
SAS>   test=t
SAS>   nullmean=10
SAS>   mean=10.50
SAS>   sides=u
SAS>   stddev=1.2
SAS>   ntotal=18
SAS>   power=.;

The POWER Procedure
One-Sample t Test for Mean
     Fixed Scenario Elements
Distribution                Normal
Method                       Exact
Number of Sides                  U
Null Mean                       10
Mean                          10.5
Standard Deviation             1.2
Total Sample Size               18
Alpha                         0.05

Computed Power
Power
0.521
```

Figure 8: Output for carbon monoxide experiment part 1

```
SAS> proc power;
SAS>   onesamplemeans
SAS>   test=t
SAS>   nullmean=10
SAS>   mean=10.50
SAS>   sides=u
SAS>   stddev=1.2
SAS>   ntotal=.
SAS>   power=0.80;

The POWER Procedure
One-Sample t Test for Mean
     Fixed Scenario Elements
Distribution                Normal
Method                       Exact
Number of Sides                  U
Null Mean                       10
Mean                          10.5
Standard Deviation             1.2
Nominal Power                  0.8
Alpha                         0.05

Computed N Total
Actual        N
 Power    Total
 0.810       38
```

Figure 9: Output for carbon monoxide experiment part 2

```
R> macaque=read.table("macaque.txt",header=T)
R> attach(macaque)
R> head(macaque)

  tone call
1  474  500
2  256  138
3  241  485
4  226  338
5  185  194
6  174  159
```

Figure 10: Macaque data for brain responses to sound (part)

```
R> d=tone-call
R> obs=mean(d)
R> obs

[1] -59.56757
```

Figure 11: Macaque analysis, part 1

```
R> shuffle=function(x) {
R>   sgn=sample(c(-1,1),length(x),replace=T)
R>   m=mean(abs(x)*sgn)
R>   return(m)
R> }
R> replicate(5,shuffle(d))

[1]  14.378378  21.351351 -22.810811  12.972973  -3.675676
```

Figure 12: Macaque analysis, part 2

```
R> rand=replicate(1000,shuffle(d))
R> hist(rand)
R> abline(v=obs,lty="dashed")
```

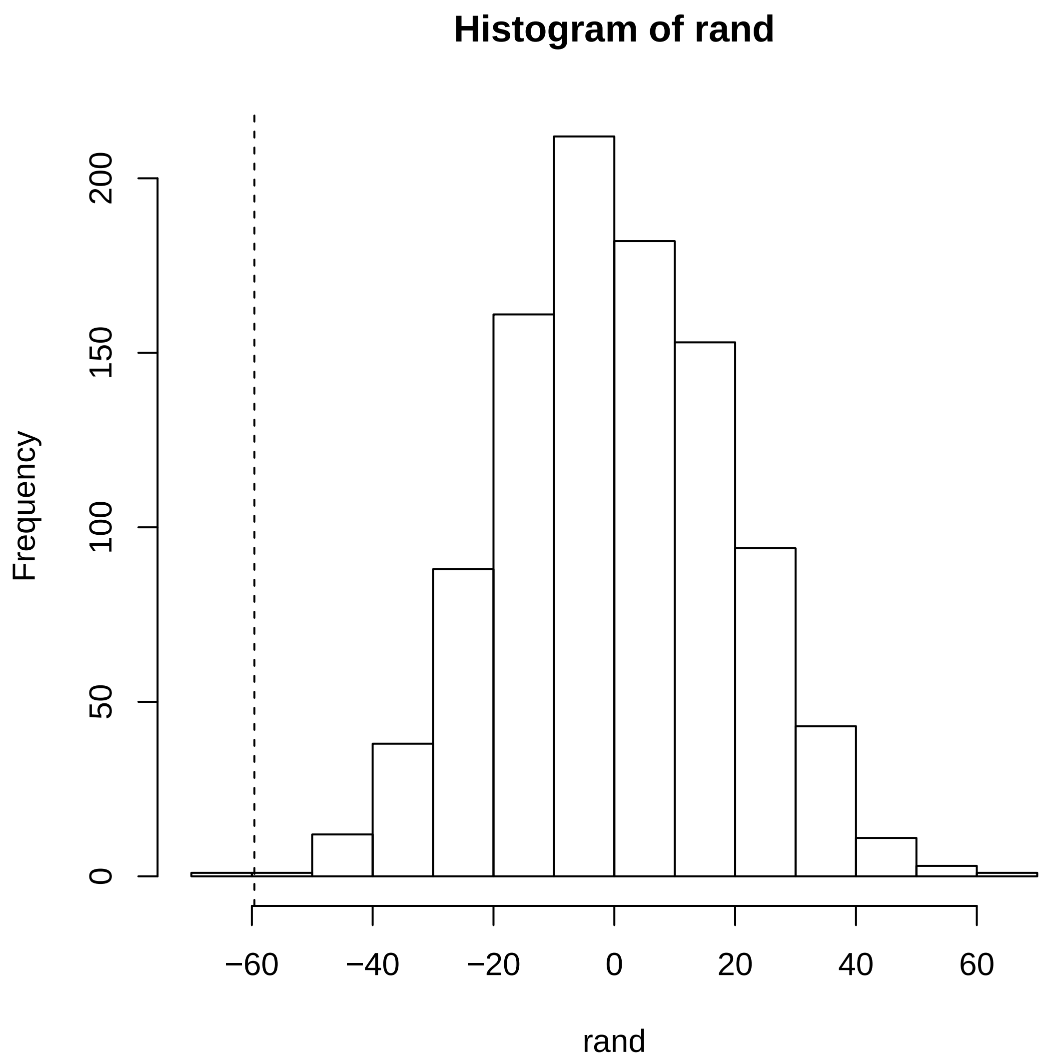

Figure 13: Macaque analysis, part 3

```
R> table(rand<obs)

FALSE  TRUE
  999     1
```

Figure 14: Macaque analysis, part 4

```
SAS>   data truck2;
SAS>     set '/home/ken/trucking';
SAS>     logpricptm=log(pricptm);
SAS>     origin01=(origin='JAX');
SAS>     dereg01=(dereg='YES');
```

Figure 15: Trucking data

```
SAS>   proc reg;
SAS>   model logpricptm=mileage shipment pctload origin01 dereg01;

The REG Procedure
Model: MODEL1
Dependent Variable: logpricptm
Number of Observations Read         448
Number of Observations Used         448

                            Analysis of Variance
                                   Sum of           Mean
Source                    DF      Squares         Square    F Value    Pr > F
Model                      5    130.72883       26.14577     134.12    <.0001
Error                    442     86.16643        0.19495
Corrected Total          447    216.89526

Root MSE              0.44153    R-Square     0.6027
Dependent Mean       10.85452    Adj R-Sq     0.5982
Coeff Var             4.06769

                        Parameter Estimates
                      Parameter       Standard
Variable      DF       Estimate          Error    t Value    Pr > |t|
Intercept      1       12.28946        0.06253     196.53      <.0001
MILEAGE        1       -0.29828        0.01515     -19.68      <.0001
SHIPMENT       1        0.22453        0.10598       2.12      0.0347
PCTLOAD        1       -0.06190        0.02543      -2.43      0.0153
origin01       1       -0.17463        0.04200      -4.16      <.0001
dereg01        1       -0.41333        0.04294      -9.63      <.0001
```

Figure 16: First regression for trucking data

```
SAS>   proc reg;
SAS>   model logpricptm=mileage origin01 dereg01;

The REG Procedure
Model: MODEL1
Dependent Variable: logpricptm
Number of Observations Read         448
Number of Observations Used         448

                             Analysis of Variance
                                    Sum of           Mean
Source                  DF         Squares         Square    F Value    Pr > F
Model                    3        91.95243       30.65081     108.92    <.0001
Error                  444       124.94283        0.28140
Corrected Total        447       216.89526

Root MSE             0.53047    R-Square     0.4239
Dependent Mean      10.85452    Adj R-Sq     0.4201
Coeff Var            4.88713

                         Parameter Estimates
                      Parameter       Standard
Variable      DF       Estimate          Error    t Value    Pr > |t|
Intercept      1       11.99321        0.07073     169.57      <.0001
MILEAGE        1       -0.30330        0.01818     -16.68      <.0001
origin01       1       -0.16724        0.05044      -3.32      0.0010
dereg01        1       -0.40231        0.05139      -7.83      <.0001
```

Figure 17: Second regression for trucking data

Berkeley	Duke	Jersey	Sierra
5.13	5.31	5.20	5.08
5.36	4.89	4.92	5.30
5.20	5.09	5.44	5.43
5.15	5.57	5.20	4.99
4.96	5.36	5.17	4.89
5.14	4.71	5.24	5.30
5.54	5.13	5.08	5.35
5.22	5.30	5.13	5.26

Figure 18: Untidy blueberry data

```
R> life %>% select(-1) %>% summary()

    lifeexp         lifeexpf        lifeexpm        logperdr     
 Min.   :51.50   Min.   :53.00   Min.   :50.00   Min.   : 5.421  
 1st Qu.:64.12   1st Qu.:65.25   1st Qu.:61.25   1st Qu.: 6.124  
 Median :70.00   Median :73.00   Median :66.50   Median : 6.700  
 Mean   :67.76   Mean   :70.37   Mean   :65.21   Mean   : 7.052  
 3rd Qu.:74.12   3rd Qu.:77.75   3rd Qu.:70.50   3rd Qu.: 7.956  
 Max.   :79.00   Max.   :82.00   Max.   :76.00   Max.   :10.509  
    logpertv     
 Min.   :-1.470  
 1st Qu.: 1.163  
 Median : 1.757  
 Mean   : 2.256  
 3rd Qu.: 3.113  
 Max.   : 6.384  

R> life %>% select(-1) %>% cor()

            lifeexp   lifeexpf   lifeexpm   logperdr   logpertv
lifeexp   1.0000000  0.9952704  0.9932558 -0.7990433 -0.6570390
lifeexpf  0.9952704  1.0000000  0.9781375 -0.8018263 -0.6631969
lifeexpm  0.9932558  0.9781375  1.0000000 -0.7755782 -0.6396490
logperdr -0.7990433 -0.8018263 -0.7755782  1.0000000  0.5591676
logpertv -0.6570390 -0.6631969 -0.6396490  0.5591676  1.0000000
```

Figure 19: Life expectancy data summaries

```
R> cutoff=2*(5+1)/38
R> z=rep(1,38)
R> z.1=lm(z~lifeexp+logpertv+logperdr+lifeexpf+lifeexpm,data=life)
R> h=hatvalues(z.1)
R> life %>% mutate(lev=h) %>% filter(lev>cutoff)

   country lifeexp lifeexpf lifeexpm  logperdr  logpertv       lev
1 Ethiopia    51.5       53       50 10.509442  6.220590 0.3956676
2    Sudan    53.0       54       52  9.437476 -1.469676 0.7256571
3 Thailand    68.5       73       66  8.493515  2.397895 1.0000000
```

Figure 20: Life expectancy data analysis

```
BrakePower Fuel            MassBurnRate
4          DF-2            13.2
4          Blended         17.5
4          AdvancedTiming  17.5
6          DF-2            26.1
6          Blended         32.7
6          AdvancedTiming  43.5
8          DF-2            25.9
8          Blended         46.3
8          AdvancedTiming  45.6
10         DF-2            30.7
10         Blended         50.8
10         AdvancedTiming  68.9
12         DF-2            32.3
12         Blended         57.1
```

Figure 21: Data for R plot

antimony	method	s1	s2	s3
0	AB	18.3	19.8	22.9
0	FC	19.4	19.8	20.3
0	OQ	20	24.3	21.9
0	WQ	17.6	19.5	18.3
3	AB	21.7	22.9	22.1
3	FC	19	20.9	19.9
3	OQ	20	20.9	20.4
3	WQ	18.6	19.5	19
5	AB	22.9	19.7	21.6
5	FC	19.6	16.4	20.5
5	OQ	20.9	22.9	20.6
5	WQ	22.3	19.5	20.5
10	AB	15.8	17.3	17.1
10	FC	16.4	17.6	17.6
10	OQ	16.4	19	18.1
10	WQ	15.2	17.1	16.6

Figure 22: Tin-lead data