Test 3 in 214 will be Thursday May 12. It will cover the following sections in the book:

8.1,2,3; 9.1,2,3,4,5; 11.1, 2,5,

We will leave the material in 11.4 and 12.2 for the final exam. (No calculations, only reading computer output and understanding what the model is.)

We did not cover: 8.4, 11.3, and 12.4

Test 3 is basically the material on significance tests. The basic setup is important to understand. The pieces are

1. A pair of hypotheses $H_0$ and $H_A$.
2. A test statistic with known sampling distribution to judge the validity of an assumption that $H_0$ is true
3. A data set that yields an observed value of the test statistic
4. An analysis computing a $p$-value, or atleast comparing it to a significance level $\alpha$.

We learned the following tests:

1. test for proportion
2. Test for mean
3. test of difference of propotions
4. test for difference of means (with different assumptions: equal variance, non-equal variance, paired data vs. independent data)
5. Test for the regression coefficients $\beta_0$ and $\beta_1$.

(in 11.4 and 12.2 we learn about two more significance tests both related to the $F$ statistic.)

Each test has a different test statistic and a different set of assumptions. However all the test statistics (except for a single proportion) we met have this form

$$\frac{\text{observed} - \text{expected}}{\mathsf{SE}}$$

What these means differs depending on the problem. For instance we have all of these possibilities for SE:

$$s/\sqrt{n}, \sqrt{\widehat{p}(1-\widehat{p})(1/n_1+1/n_2)}, s_p\sqrt{1/n_1+1/n_2}, \sqrt{s_1^2/n_1+s_2^2/n_2}, \ldots$$

# 1 Problemos

For each significance test you should do all of the following:

1. Specify the null and alternative hypotheses. This is really important, as once you've written this down, the choice of a test statistic follows and after that the problem becomes mechanical.
2. Ask your self if the assumptions that allow you to "know" the sampling distribution of the test statistic are met. (I.e., normality of the population(s), variances are equivalent, large sample, ...)
3. Find the $p$-value (or atleast find bounds for it) and compare it to $\alpha = 0.05$. Answer if the differences are statistically significant. As well, answer if you *accept* or *reject* the null hypothesis.

Additionally, if you can, draw a nice picture of the sampling distribution of the test statistic, and shade in either the $p$-value or the critical values so that you can see the answer.

1. The English mathematican John Kerrich tossed a coin 10,000 times and obtained 5,06 heads. Is this significant evidence (at the 5% level) that the probability that Kerrich's coin comes up heads is different from 0.5?
2. Land's Beginning is considering buying a mailing list of 100,000 people, but only if they are reasonably confident that 5% (or more) will respond to direct mailing. They are given a random sample of 500 from the list and find that only 4% responded to a mailing. Is this difference from 5% statistically significant to make LB not want to buy these names?
3. Christmas tree growers want to know if there is a difference in preference for natural vs. artificial trees between city and suburb dwellers among those who buy christmas trees. They perform a small survey and find these preferences:

   ```
   population n prefer natural
   ---------------------------
   urban      160        64
   suburban   261        89
   ```

   Is the difference statistically significant?

4. Does removing a annual credit card fee change consumer purchases? A bank investigates by waiving the fee on a random sample of 500 of its customers and tracks differences in the difference in the amount they spent. The mean increase was \$565 with $s = \$267$. Is there significant evidence at the 5% level that the mean amount increased?

5. How accurate are home radon detectors? A company tested 6 in an environment with an *accurately* measured amount of 115. The 6 cheap detectors found

```
91.9  122.3 105.4  95.0 99.6  120.9
```

   with summary

```
     xbar        sd         n
105.85000  13.02839   6.00000
```

   Is there evidence to indicate that the home measurements are lower than they should be?

6. Do piano lessons improve spatio-temporal reasoning in children? To test, two independent groups were formed, one given piano lessons for one month the other not. Differences in reasoning were assessed numerically and are summarized by

```
group       n     xbar    s
----------------------------
piano      12     10.7    3.8
control    11      9.6    4.0
```

   Is the difference statistically significant? (Did you assume equal variances? independent or paired samples?)

7. A study of iron deficiency between children who were bottle-fed versus nursed is summarized by

```
group     n    xbar    s
-------------------------
Nursed   23    13.3   1.7
Bottle   19    12.4   1.8
```

Is the difference statistically significant? (Did you assume equal variances? independent or paired samples?)

8. Does cocaine use by expectant mothers cause smaller birthweights? A restrospective study presents the following data

| Group | n | xbar | s |
|---|---|---|---|
| Cocaine | 134 | 2733 | 599 |
| No Cocaine | 5974 | 3118 | 672 |

Is the difference statistically significant? (Did you assume equal variances? independent or paired samples?)

9. A classic test on the effect of a home environment on IQ was performed in the 60s by Burt by using pairs of identical twins that were separated at birth. One twin raised by foster parents, one raised by the birth parents. For each pair, IQ scores were recorded. Suppose the data recorded was

| | data | | | | | n | xbar | s |
|---|---|---|---|---|---|---|---|---|
| foster | 120 | 101 | 95 | 97 | 97 | 5 | 102 | 10.2 |
| biological | 129 | 102 | 97 | 99 | 105 | 5 | 106.4 | 13 |

Let $\mu_f$ and $\mu_b$ be the population means for all such twins. Perform a two-sided significance test of $H_0 : \mu_f = \mu_b$.

(a) What is the $p$ value?

(b) Why did the researcher used twins, and not two random samples?

(c) What assumptions did you make about the data?

(d) What would it mean if the null hypothesis was true? (Why is this a test of the effect of environment?)

10. A data set containing information about various factors effect on crime shows the following relationship between the state GDP and probability of incarceration for 47 states. A linear model is fit and is illustrated below

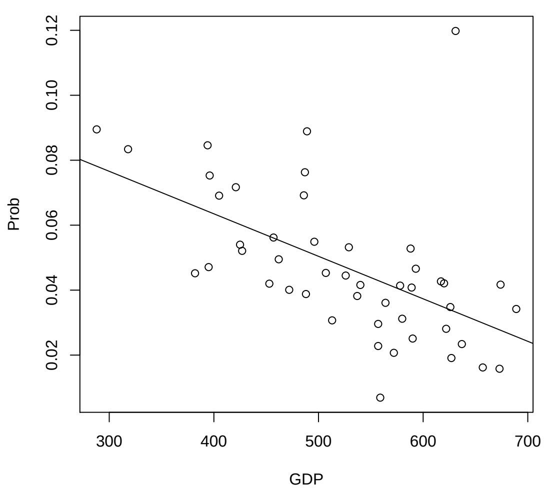
```
> res = lm(Prob ~ GDP, UScrime)
> plot(Prob ~ GDP, UScrime)
> abline(res)
> summary(res)

Call:
lm(formula = Prob ~ GDP, data = UScrime)

Residuals:
      Min        1Q    Median        3Q       Max
-0.035792 -0.013269 -0.003581  0.008438  0.086533

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.158e-01  1.560e-02   7.426 2.39e-09 ***
GDP         -1.309e-04  2.921e-05  -4.480 5.09e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01912 on 45 degrees of freedom
Multiple R-Squared: 0.3084,      Adjusted R-squared: 0.293
F-statistic: 20.07 on 1 and 45 DF,  p-value: 5.086e-05
```



(a) Write down the regression line, and make a prediction for a missing state with GDP of 600.

(b) Is a significance test of

$$H_0 : \beta_1 = 0, \quad H_A : \beta_1 \neq 0$$

Accepted or rejected at the 0.05 level?

(c) Circle any points that are outliers for the regression model.

(d) Does it really make sense to do statistical inference in this example, as the data is basically census data?

11. Does the number of beers a student drinks predict his or her blood alcohol level? A data set from Ohio State University of 16 students is summarized below

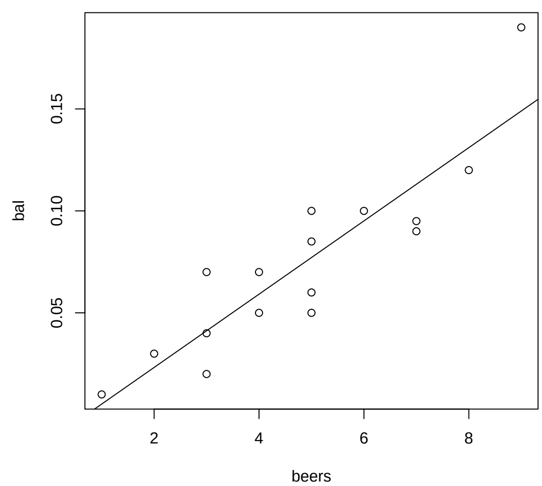
```
> beers = c(5, 2, 9, 8, 3, 7, 3, 5, 3, 5, 4, 6, 5, 7, 1, 4)
> bal = c(0.1, 0.03, 0.19, 0.12, 0.04, 0.095, 0.07, 0.06, 0.02,
+     0.05, 0.07, 0.1, 0.085, 0.09, 0.01, 0.05)
> res = lm(bal ~ beers)
> plot(bal ~ beers)
> abline(res)
> summary(res)

Call:
lm(formula = bal ~ beers)

Residuals:
      Min        1Q    Median        3Q       Max
-0.027118 -0.017350  0.001773  0.008623  0.041027

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.012701   0.012638  -1.005    0.332
beers        0.017964   0.002402   7.480 2.97e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02044 on 14 degrees of freedom
Multiple R-Squared: 0.7998,      Adjusted R-squared: 0.7855
F-statistic: 55.94 on 1 and 14 DF,  p-value: 2.969e-06
```




(a) What is the predicted BAL for a student who drinks 5 beers?

(b) The students were not controlled for weight or gender. Does the value of $r^2$ reflect this? Explain.

(c) Test the hypothesis

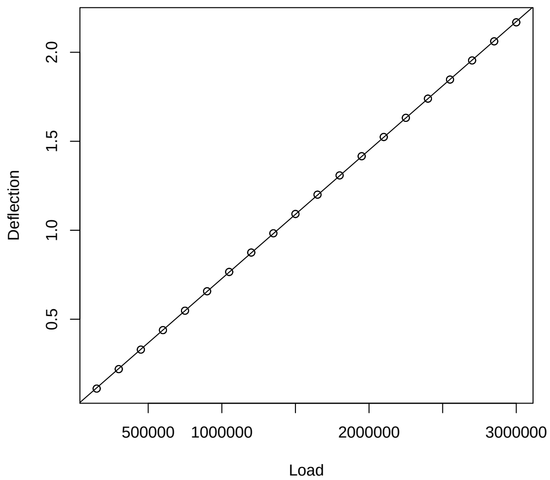
$$H_0 : \beta_0 = 0, \quad H_A : \beta_0 \neq 0$$

(d) A rule of thumb says that 5 beers will bring one to the legal limit. Suppose this is 0.08, so that the slope would be expected to be $0.08/5 = 0.016$. Test the hypothesis

$$H_0 : \beta_1 = 0.016, \quad H_A : \beta_1 \neq 0.016.$$

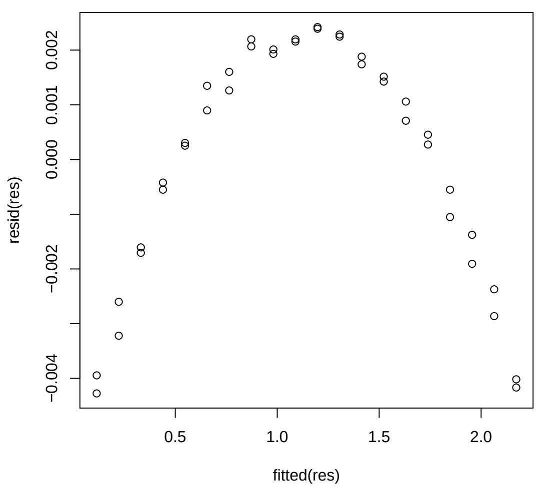
12. The scatterplot shows a linear model fitting a deflection for a given load amount (how much something changes based on a weight placed on it).

```
> plot(Deflection ~ Load, deflection)
> res = lm(Deflection ~ Load, deflection)
> abline(res)
```




A Residual plot shows

```
> plot(resid(res) ~ fitted(res))
```




(a) Is this model appropriate for linear regression analysis?

(b) Does the high value of $R^2$ (it rounds to 1 to two decimal points) indicate that a linear model is a good fit for the data?