Chapter 2

Inferences for Simple Linear Regression

2.1 Hypothesis Tests for ρ, β_1, and β_0

Assuming i.i.d. normal errors and fixed explanatory variable values, the slope estimator b_1 has a normal distribution with mean and variance

$$E[b_1] = \beta_1, \qquad \mathrm{Var}(b_1) = \frac{\sigma^2}{\sum_i (X_i - \overline{X})^2}.$$

Refer to ALSM for details. Thus

$$\frac{b_1 - \beta_1}{\sqrt{\mathrm{Var}(b_1)}} = \frac{b_1 - \beta_1}{\sigma} \sqrt{\sum_i (X_i - \overline{X})^2}$$

behaves according to a standard normal distribution. However, we don't know σ, so we estimate it with $\sqrt{MSE}$ (since MSE is an unbiased estimator for σ^2). So

$$\boxed{\frac{b_1 - \beta_1}{\sqrt{\frac{MSE}{\sum_i (X_i - \overline{X})^2}}} \sim t_{(n-2)}.} \tag{2.1}$$

This is all we need to make CIs and hypothesis tests for β_1. [1]

[1] Setting $\beta_1 = 0$ in (2.1) gives

$$b_1 \sqrt{\frac{\sum_i (X_i - \overline{X})^2}{MSE}} = r \frac{S_Y}{S_X} \sqrt{\frac{\sum_i (X_i - \overline{X})^2}{MSE}} = r S_Y \sqrt{\frac{n-1}{MSE}} = r \sqrt{\frac{n-2}{1-r^2}} \sim t_{n-2}$$

so that testing $H_0 : \beta_1 = 0$ is the same as testing $H_0 : \rho = 0$ using $r\sqrt{\frac{n-2}{1-r^2}}$ as a test statistic.

Under the same assumptions (i.i.d. normal errors and fixed explanatory values),

$$E[b_0] = \beta_0, \qquad \text{Var}(b_0) = \sigma^2 \left(\frac{1}{n} + \frac{\overline{X}^2}{\sum_i (X_i - \overline{X})^2} \right)$$

and so

$$\frac{b_0 - \beta_0}{\sqrt{\text{Var}(b_0)}} = \frac{b_0 - \beta_0}{\sigma \left(\frac{1}{n} + \frac{\overline{X}^2}{\sum_i (X_i - \overline{X})^2} \right)^{1/2}}$$

behaves as a standard normal, so that

$$\boxed{\frac{b_0 - \beta_0}{\sqrt{MSE} \left(\frac{1}{n} + \frac{\overline{X}^2}{\sum_i (X_i - \overline{X})^2} \right)^{1/2}} \sim t_{n-2}.}$$

This is all we need to make CIs and hypothesis tests for β_0.

It turns out these procedures still work pretty well even if the response variables aren't exactly normal. Also, even if they are far from normal, as long as the sample sizes are large, the distributions of b_1 and b_0 are approximately the same as above.

Example 6 *Used Cars.* Make 95% CIs for the actual values of β_0 and β_1 for the original ten data values. Also, test if there is significant statistical evidence to suggest that $\beta_1 < 0$ and $\beta_0 \neq 0$.

By "hand":

```
> age <- c(3, 6, 8, 4, 2, 11, 4, 7, 7, 9)
> price <- c(172, 140, 112, 160, 165, 80, 155, 103, 84, 78)
> b1 <- cor(age, price) * sd(price) / sd(age)
> b0 <- mean(price) - b1* mean(age)
> b1

[1] -12.15226

> b0

[1] 199.0288

> fits <- b1*age + b0
> residuals <- price - fits
> SSE <- sum(residuals^2)
> MSE <- SSE / (10-2)
> MSE
```



```
[1] 245.1512

> sqrt(MSE)

[1] 15.65731

> stdev.b1 <- sqrt(MSE/sum((age - mean(age))^2))
> stdev.b0 <- sqrt(MSE) * sqrt(1/10 + (mean(age) ^2)/ sum((age - mean(age))^2))
> stdev.b1

[1] 1.833806

> stdev.b0

[1] 12.23301

> #########  95% CI for beta_1
> left.end.pt <- b1 - qt(.975, 8)*stdev.b1
> right.end.pt <- b1 + qt(.975, 8)*stdev.b1
> left.end.pt

[1] -16.38103

> right.end.pt

[1] -7.9235

> #########  95% CI for beta_0
> left.end.pt <- b0 - qt(.975, 8)*stdev.b0
> right.end.pt <- b0 + qt(.975, 8)*stdev.b0
> left.end.pt

[1] 170.8194

> right.end.pt

[1] 227.2382
```

```
> #########  Hypo test: H_a: beta_1 < -10
> #########  Perhaps cars depreciated at rate
> #########  $1000 per year in the past, and
> #########  so we want to know if there's much
> #########  statistical evidence that the depreciation
> #########  rate has increased.
> ts.beta1 <- (b1 - (-10))/stdev.b1
> ts.beta1

[1] -1.17366

> p.value <- pt( -1.17366, 8)
> p.value

[1] 0.1371432

> #############   Or...
> #############  H_a: beta_1 differs from 0....
> ts.beta1 <- (b1 - 0)/stdev.b1
> p.value <- 2*pt(-abs((b1-0)/stdev.b1), 8)
> p.value

[1] 0.0001647268

#############  H_a: beta_0 differs from 0...
> p.value <- 2*pt(-abs((b0-0)/stdev.b0), 8)
> p.value

[1] 2.049246e-07
```

With R:

```
> summary(regression)

Call:
lm(formula = used_cars$price ~ used_cars$age)

Residuals:
    Min      1Q  Median      3Q     Max
-29.963 -10.653   7.004  10.037  14.646

Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    199.029     12.233  16.270 2.05e-07 ***
used_cars$age  -12.152      1.834  -6.627 0.000165 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 15.66 on 8 degrees of freedom
Multiple R-squared:  0.8459,    Adjusted R-squared:  0.8266
F-statistic: 43.91 on 1 and 8 DF,  p-value: 0.0001647
```


2.2 Power

The *power* of a hypothesis test is the probability of avoiding type II error, given the alternative hypothesis is true for some particular value of the parameter in question. You can think of it as the probability that you will detect a difference when there really is one. Power calculations always require type I error probabilities (α) because they involve the decision to reject or fail to reject H_0, and this decision is based comparing the p-value to α (or equivalently the test statistic value to a critical value).

Example 7 *Used Cars.* Suppose β_1 is in fact -5 (indicating cars actually tend to drop in value at a rate of $500 per year) and we set $\alpha = 0.05$. Then in testing

$$\begin{aligned} H_0 : &\quad \beta_1 = 0 \\ H_0 : &\quad \beta_1 < 0 \end{aligned}$$

note that

$$\frac{b_1 - (-5)}{s(b_1)} \sim t_{\text{d.f.}=8}.$$

Suppose further that based on past experience we think that for these ages, the standard deviation of $b_1 = \hat{\beta}_1$ is thought to be about 2. Then

$$\begin{aligned} \text{power} &= P\left(\frac{b_1}{2} < t(.05, 8) \Big| \frac{b_1 + 5}{2} \sim t_{d.f.=8}\right) \\ &= P\left(\frac{b_1 + 5}{2} < t(.05, 8) + 5/2 \Big| \frac{b_1 + 5}{2} \sim t_{d.f.=8}\right) \\ &= \text{pt(2.5 + qt(.05, 8), 8)} \\ &= 0.7301. \end{aligned}$$

You should see the power increases as the difference between the actual value of β_1 and the one claimed in the null hypothesis increases. For example, if β_1 is actually -8, the power is 0.9676346, and when β_1 is actually -20, the power is 0.9999807.

2.3 CIs for Mean Responses

Suppose we fix the the values of the explanatory variable (say at $x_1, x_2, ...$, and note they can, but do not have to be, unique). Then we measure the responses and get a least-squares regression line. Then do this all over again many times for the same set of $x - values$:

Then maybe we're interested in obtaining a CI for the predicted response $\hat{Y}_h$ for one fixed x value (call it x_h). It turns out $\hat{Y}_h$ behaves according to a normal distribution with mean and variance

$$E[\hat{Y}_h|X = x_h] = E[Y|X = x_h], \qquad \text{Var}(\hat{Y}_h|X = x_h) = \sigma^2 \left[\frac{1}{n} + \frac{(x_h - \overline{x})^2}{\sum_i (x_i - \overline{x})^2}\right].$$

As before, we use MSE to estimate the common variance σ^2 of the errors, so that

$$\widehat{\text{Var}(\hat{Y}_h|X} = x_h) = MSE\left[\frac{1}{n} + \frac{(x_h - \overline{x})^2}{\sum_i (x_i - \overline{x})^2}\right].$$

Note that setting $x = 0$ in the variance expression above gives the variance of $b_0 = \hat{\beta}_0$.

Then

$$\frac{\hat{Y}_h - E[Y|X = x_h]}{\sqrt{MSE\left[\frac{1}{n} + \frac{(x_h - \overline{x})^2}{\sum_i (x_i - \overline{x})^2}\right]}} \sim t \text{ with } n - 2 \text{ degrees of freedom}$$

and we can be $(1 - \alpha) \times 100\%$ confident that the actual value of $E[\hat{Y}_h|X = x_h]$ is within the interval with endpoints

$$\hat{y}_h \pm t(1 - \alpha/2, n - 2) \times \sqrt{MSE\left[\frac{1}{n} + \frac{(x_h - \overline{x})^2}{\sum_i (x_i - \overline{x})^2}\right]}.$$

Example 8 *Used Cars.* Use the data to find a 98% confidence interval for $E[\hat{Y}|X = 4]$.

By "hand" with R:

```
> b1 <- cor(age, price)* sd(price)/sd(age)
> b0 <- mean(price)-b1*mean(age)
> fits <- b1*age + b0
> residuals <- price - fits
> SSE <- sum(residuals^2)
> MSE <- SSE /8
> sterror <- sqrt(MSE*(.10 + ((4-mean(age))^2)/sum((age - mean(age))^2)))
> lpt <- 4*b1+b0 - qt(.975, 8) * sterror  ## 95% CI
> rpt <- 4*b1+b0 + qt(.975, 8) * sterror  ## 95% CI
> lpt

[1] 135.9552

> rpt

[1] 164.8843

> lpt <- 4*b1+b0 - qt(.99, 8) * sterror ##### 98% CI...
> rpt <- 4*b1+b0 + qt(.99, 8) * sterror ##### 98% CI...
> lpt

[1] 132.2515

> rpt

[1] 168.588
```

With R:

```
> age <- c(3, 6, 8, 4, 2, 11, 4, 7, 7, 9)
> price <- c(172, 140, 112, 160, 165, 80, 155, 103, 84, 78)
> data <- data.frame(price, age)
> attach(data)
> out <- lm(data)
> nd = data.frame(age=4)
> predict(out, nd, interval="confidence", level=.95)
```



```
       fit      lwr      upr
1 150.4198 135.9552 164.8843
```

```
> predict(out, nd, interval="confidence", level=.98)
```

```
       fit      lwr     upr
1 150.4198 132.2515 168.588
```

A few things:

1. Decreasing in MSE results in a smaller interval. However, we often have little control over MSE, other than including additional explanatory variables which might account for some of this error, or, measuring the response variable values as accurately as possible.
2. Increasing the confidence level increases the size of the interval (as usual).
3. Increasing the sample size n tends to cause the interval to decrease in width.
4. Narrower CIs can be achieved by taking values of the predictor variable that are more spread out.
5. The CI will tend to be smaller in width if x_h is chosen near the mean of the predictor variable values.
6. Only use the method for making a CI for EY_h if x_h is between the minimum and maximum values of the observed predictor variable values.
7. You should only use this method for making a CI for EY_h when the model conditions are met: linearity, independent normally distributed errors with mean 0 and common variance σ^2.

2.4 Prediction Intervals for New Responses

If the regression parameters β_0, β_1, and σ are known, under the typical model assumptions (normality, etc.), the mean response corresponding to predictor variable value of $X = x_h$ is

$$E[Y_h|X = x_h] = \beta_0 + \beta_1 x_h.$$

The behavior of Y_h is normal. Then a $(1-\alpha) \times 100\%$ prediction interval (different from a confidence interval) is an interval which contains the response with probability $1-\alpha$. An easy way to get this interval is to use a symmetric one about the point estimate (since normal pdfs are symmetric). If the regression parameters β_0, β_1, and σ are unknkown (which is likely the case), the formula is considerable more difficult: if $X = x_h$, there is a probability of about $1-\alpha$ that the response will occur between

$$\hat{y_h} \pm t(1-\alpha/2, n-2) \times \sqrt{MSE\left(1 + \frac{1}{n} + \frac{(x_h - \overline{x})^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\right)}$$

Example 9 *Used Cars.* Make 95% and 98% prediction intervals for the price for a 4 year old car.
By "hand":

```
> pred.sterror <- sqrt(MSE*(1+
+ .10 + ((4-mean(age))^2)/sum((age - mean(age))^2)))
> lpt <- 4*b1+b0 - qt(.975, 8) * pred.sterror
> rpt <- 4*b1+b0 + qt(.975, 8) * pred.sterror
> lpt

[1] 111.5243

> rpt

[1] 189.3152

> lpt <- 4*b1+b0 - qt(.99, 8) * pred.sterror
> rpt <- 4*b1+b0 + qt(.99, 8) * pred.sterror
> lpt

[1] 101.5651

> rpt
```



```
[1] 199.2744
```

With R:

```
> nd = data.frame(age=4)
> predict(out, nd, interval="prediction", level=.95)


       fit      lwr      upr
1 150.4198 111.5243 189.3152


> predict(out, nd, interval="prediction", level=.98)


       fit      lwr      upr
1 150.4198 101.5651 199.2744
```

You should only use this procedure when x_h is between the minimum and maximum values observed for the predictor variable, and also when all the regular regression assumptions are met (linearity and independent normally distributed errors with mean 0 and common variance σ^2).

2.5 Confidence Bands

Use the procedures in the previous sections, but for each predictor variable value between the minimum and maximum predictor variable values.

Example 10 Used Cars.

```
> x<- age
> y<- price
> d <- data.frame(x, y)
> lm1 <- lm(y~x,data=d)
> p_conf1 <- predict(lm1,interval="confidence", level=.95)
> p_pred1 <- predict(lm1,interval="prediction", level=.95)
> nd <- data.frame(x=seq(0,12,length=51))
> p_conf2 <- predict(lm1,interval="confidence",newdata=nd, level = .95)
> p_pred2 <- predict(lm1,interval="prediction",newdata=nd, level = .95)
> plot(y~x,data=d,ylim=c(0,250),xlim=c(-1,14))
> abline(lm1)
> matlines(nd$x,p_conf2[,c("lwr","upr")],col=4,lty=1,type="b",pch="c")
> matlines(nd$x,p_pred2[,c("lwr","upr")],col=3,lty=2,type="b",pch="p")
```

The Working-Hotelling procedure gives confidence bands for *the entire regression line at once*, with boundary values given by

$$b_0 + b_1 x \pm W\sqrt{MSE}\left[\frac{1}{n} + \frac{(x-\overline{x})^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}\right]^{1/2},$$

where

$$W^2 = 2F(1-\alpha; 2, n-2).$$

Note the similarity between this formula and the formula for CIs for mean response: they are the *same* but with a W-score instead of a t-score. The W-factor is larger than the t-factor because we're attempting to surround the entire least squares regression line. The interval end points for the Working-Hotelling get farther apart as X_h gets farther from the mean of the x-values.

Example 11 *Used Cars.* Make 95% Working-Hotelling CI bands for the regression line.

```
> W <- sqrt(2 * qf(.95, 2, 8))
> WHbands <- p_conf2  # since the CIs for means are the same save the F statistic
> WHL <- (WHbands[,2]+WHbands[,3])/2 - (WHbands[,3] - WHbands[,2])/2/qt(.975, 8)*W
> WHH <- (WHbands[,2]+WHbands[,3])/2 + (WHbands[,3] - WHbands[,2])/2/qt(.975, 8)*W
> WHbands <- cbind(WHbands, WHL, WHH)
> matlines(nd$x,WHbands[,c("WHL","WHH")],col=5,lty=1,type="b",pch="*")
```


Or you can do it like this:

```
> CI <- predict(lm1, se.fit=TRUE)
> W <- sqrt(2*qf(0.95,length(lm1$coefficients), lm1$df.residual))
> Band <- cbind(CI$fit - W * CI$se.fit, CI$fit + W * CI$se.fit )
> points(sort(used_cars$age), sort(Band[,1], decreasing=TRUE), type="l", lty=2)
> points(sort(used_cars$age), sort(Band[,2], decreasing=TRUE), type="l", lty=2)
> legend("topright",legend=c("Mean Line","95% CB"),col=c(1,1),lty=c(1,2),bg="gray90")
```

If the regression line had a positive slope, we'd need to remove the "decreasing=TRUE" or change to "FALSE" in the sort argument.

2.6 Simultaneous Inferences

We often want to make several estimates (called a *family* of estimates) at the same time. For instance, we might want to make CIs for β_0 and β_1, or perhaps construct CIs for several predicted values or mean responses. We want make a statement like, "we can be at least ??% confident that all of these intervals contain the respective parameters."

The *Bonferroni inequality* says for any events $E_1, E_2, \ldots, E_n$,

$$P(\bigcap_{i=1}^{n} E_i) \geq \sum_{i=1}^{n} P(E_i) - n + 1.$$

We can apply this inequality to make a family of n CIs for various model parameters by considering E_i to be the event that the ith random interval contains the ith parameter. Also, maybe we could set all the the individual, or marginal, probabilities to all be the same (say, $1-\alpha$... although this is not necessary). Then there is *at least a* $(1-n\alpha)\%$ chance that all n CIs catch their respective target parameters. Note this means the individual type I error rate α is always at least the family type I error rate α_f/n.

Example 12 *Used Cars.* Make simultaneous Bonferroni CIs for the slope and intercept with a family confidence level of 0.98.

```
age <- c(3, 6, 8, 4, 2, 11, 4, 7, 7, 9)
price <- c(172, 140, 112, 160, 165, 80, 155, 103, 84, 78)
b1 <- cor(age, price) * sd(price) / sd(age)
b0 <- mean(price) - b1* mean(age)
fits <- b1*age + b0
residuals <- price - fits
SSE <- sum(residuals^2)
MSE <- SSE / (10-2)
stdev.b1 <- sqrt(MSE/sum((age - mean(age))^2))
stdev.b0 <- sqrt(MSE) * sqrt(1/10 + (mean(age) ^2)/ sum((age - mean(age))^2))
b1left.end.pt <- b1 - qt(.995, 8)*stdev.b1
b1right.end.pt <- b1 + qt(.995, 8)*stdev.b1
################### CI for beta_1
b1left.end.pt
b1right.end.pt
b0left.end.pt <- b0 - qt(.995, 8)*stdev.b0
b0right.end.pt <- b0 + qt(.995, 8)*stdev.b0
################## CI for beta_0
b0left.end.pt
b0right.end.pt
```

Example 13 *Used Cars.* Make simultaneous Bonferroni CIs for the expected sales prices for the ages of 4, 5, 6, 7, 8 years.

```
> data <- data.frame(price, age)
> attach(data)
> out <- lm(data)
> nd = data.frame(age=c(4, 5, 6, 7, 8))
> predict(out, nd, interval="confidence", level=.99)
```


2.7 ANOVA Approach to Regression

Recall the sums of squares $SSTO$, SSR, and SSE from 1.8), (1.9), and (1.10):

$$
\begin{aligned}
SSTO &:= \sum_{i=1}^{n}(y_i - \overline{y})^2, && (2.2)\\
SSR &:= \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2, && (2.3)\\
SSE &:= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2. && (2.4)
\end{aligned}
$$

Proposition 14 *The total sum of squares partitions as a sum of the regression sum of squares and the error sum of squares:*

$$SSTO = SSR + SSE.$$

The degrees of freedom associated with each sum of squares is

$$
\begin{aligned}
SSTO : df &= n - 1, && (2.5)\\
SSR : df &= 1, && (2.6)\\
SSE : df &= n - 2. && (2.7)
\end{aligned}
$$

Note the degrees of freedom for SSE and SSR sum to the degrees of freedom for $SSTO$.

Dividing a sum of squared errors by its degrees of freedom gives a *mean square*, which you can think of as a variance:

The *error mean square* MSE is

$$MSE := \frac{SSE}{n-2}.$$

The *regression mean square* MSR is

$$MSR := \frac{SSR}{1}.$$

The *total mean square* $MSTO$ is just the overall variance of the y_i values when you disregard the predictor variable:

$$MSTO := \frac{SSTO}{n-1}.$$

Remark 15 Although the total sum of squares partitions nicely as the sum of SSE and SSR, and the degrees of freedom for SSE and SSR sum to the degrees of freedom for $SSTO$, the mean squares do not sum in this way:

$$MSTO \neq MSR + MSE.$$

Example 16 *Used Cars.* ANOVA table with R:

```
> anova(out)


Analysis of Variance Table

Response: price
          Df  Sum Sq Mean Sq F value    Pr(>F)
age        1 10765.7 10765.7  43.914 0.0001647 ***
Residuals  8  1961.2   245.2
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

It turns out that MSE is an unbiased estimator for the common error variance, and MSR is a *biased* estimate for the common error variance:

$$\begin{aligned} E[MSE] &= \sigma^2, \\ E[MSR] &= \sigma^2 + \beta_1^2 \sum_{i=1}^{n} (x_i - \overline{x})^2. \end{aligned}$$

Note that

1. $E[MSE] = \sigma^2$ regardless of whether or not there exists a linear relationship between the predictor and response;

2. $\beta_1 = 0$ implies MSR is also an unbiased estimator for σ^2. So when $\beta_1 = 0$, we would expect MSE and MSR to be the same. Otherwise, MSR should tend to fall to the right of MSE.

So we should be able to use the fraction MSR/MSE to judge if $\beta_1 \neq 0$. It turns out if indeed $\beta_1 = 0$, then MSR/MSE has an F distribution with 1 numerator degree of freedom and $n - 2$ denominator degrees of freedom.

Example 17 *Used Cars.* With R:

```
> summary(out)
```

You can see the value of the F-statistic is 43.91 on 1 numerator degree of freedom and 8 denominator degrees of freedom. The corresponding p-value is printed in the output, but you can also calculate to be

```
> 1-pf(43.91, 1, 8)
```

```
[1] 0.0001647848
```

This test is equivalent to the t-test we performed earlier: note the p-values for the F and t tests for β_1 are indeed the same! That's because the square of a t-random variable with d degrees of freedom has the same distribution as an F random variable with 1 numerator and d denominator degrees of freedom. The F-test is only appropriate for testing if β_1 differs from 0. So if your alternative hypothesis is $H_a : \beta_1 \neq 0$, you can use either a t- or F-test. To do a one-sided test (left- or right-tailed) you should run a t-test.

Below is the general ANOVA table, which is the typical output format when using software. Note R includes the p-value for the F test, and does not include the "Total" row:

Source of Variation	DF	SS	MS	F
Regression	1	$SSR = \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2$	$MSR = \frac{SSR}{1}$	$F^* = \frac{MSR}{MSE}$
Residual Error	$n-2$	$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	$MSE = \frac{SSE}{n-2}$	
Total	$n-1$	$SSTO = \sum_{i=1}^{n}(y_i - \overline{y})^2$		

Recall from earlier that testing for β_1 is the same as testing for correlation ρ. If one of the variables can be identified as the response and the other as the explanatory variable, then you should think and communicate in terms of testing for β_1. On the other hand, if you're merely concerned about correlation and not necessarily about which variable affects the other, you should think and communicate in terms of testing for ρ.

2.8 Homework

1. A college admissions director selected 120 students at random to see if ACT score (X) could be used to predict GPA (Y) at the end of their first year. Open the data set CH01PR19 from ALSM.

   (a) Use R to construct a 99% CI for the slope coefficient β_1. Interpret this CI in a complete sentence, "We can be about 99% confident that...". Does the interval include 0? What is significant about the CI containing 0?

   (b) Run a hypothesis test (with R) to see if a linear association exists between ACT score and first year GPA. State the p-value and your conclusion.

   (c) Use R to construct a 98% CI for the intercept β_0. Interpret this CI in a complete sentence. Why is the intercept of interest? That is, what does β_0 tell you about first year GPA and ACT?

   (d) Use R to construct a 95% interval estimate of the mean freshman GPA for students whose ACT score is 29. Interpret this interval in a complete sentence.

   (e) Billy Billingsley scored a 29 on the ACT. Predict his freshman GPA with a 95% CI, and be sure to interpret the interval in a complete sentence. Use R of course.

   (f) Is the prediction interval you obtained in part (f) wider or skinnier than the interval you obtained in part (d)? Should it be? Why?

   (g) Add 95% confidence interval bands for the mean response to your plot. What is the CI for mean response when $X_h = 29$?. Also add Working-Hotelling bands for the regression line to your plot. What do these bands mean?

2. Why is the t-test more versatile than the F-test when testing hypotheses about β_1? Also, why is the F-test a one-tailed test even though the alternative hypothesis is $\beta_1 \neq 0$?

3. Prove the Bonferroni inequality: for any events $E_1, E_2, \ldots, E_n$,

$$P\left(\bigcap_{i=1}^{n} E_i\right) \geq \sum_{i=1}^{n} P(E_i) - n + 1.$$

4. Use the Bonferroni simultaneous inference approach to make CIs for the slope and intercept of the regression line for the Toluca Company data set CH01TA01 from ALSM. Use a family error rate of 5% (or family confidence level of 95%).

5. Problem 2.8 from *Applied Linear Statistical Models*, by Kutner et. al: Refer to the Toluca Company example and data. A consultant advises an

increase in the amount of one lot size unit requires an increase of about 3 in the expected number of work hours for the production item.

(a) Test to see if the increase in expected number of work hours equals or differs from the consultant's opinion.

(b) Calculate the power of your test in the previous part if in fact the consultant's recommendation is 1/2 hour too low. Assume the standard deviation of the slope coefficient to be 0.35.

(c) Why is the value of the F-statistic 105.88 given in the R output irrelevant in part (a)?

(d) Repeat the power calculation when the amount the standard is actually exceeded is 1 hour and 1.5 hours. Do these power calculations seem correct? Why?

6. It turns out you can use both the Bonferroni and the Working-Hotelling approaches to make simultaneous CIs for mean response values. Sometimes the Bonferroni CIs will be tighter than the Working-Hotelling intervals, and sometimes not, but the Bonferroni CIs tend to be larger than the Working-Hotelling ones when families are large.

   So make both kinds of simultaneous CIs the mean work hours responses for the three lot size levels of 30, 60, and 100 for the Toluca Company data set CH01TA01 (you can read about this data set in your book btw). For the Bonferroni intervals, use a family error rate of 10% (same as family confidence level of 90%). Use 90% for the confidence level for the Working-Hotelling bands. Compare the Bonferroni intervals with the Working-Hotelling ones.

7. The height of a tree is often predicted based on its species and diameter at breast height. The data set whitespruce.txt contains the breast height diameters (in centimeters) and actual heights (in meters) for 36 white spruce trees. Does evidence suggest there exists a strong linear association between breast height diameter and tree height? Justify your response by making and remarking on a fitted line plot and ANOVA F-test. Be sure to state the null and alternative hypotheses and your conclusion.

