Stat 20 Fall 2003, Quiz 3 Answers

B. M. Bolstad, bolstad@stat.berkeley.edu

Nov 24, 2003

Question 1. (9 points)

A biologist is doing research on how a particular growth hormone affects weight gain in mice. She takes 39 7 day old mice and randomly chooses 17 of the mice to receive the growth hormone and the remaining 22 mice will receive no growth hormone. Both groups of mice are fed the same diet. After 14 days she weighs each of the mice and records their weights. The group of mice who received the growth hormone had mean weight gain of 26.3g and standard deviation 3.3g. The control group had mean weight gain of 24.4g and standard deviation 2.8. You may assume the homoskedastic two sample normal model.

(a) Compute the pooled standard deviation.

Answer:

$$\begin{aligned} s_p &= \sqrt{\frac{(n_1 - 1)\, s_2^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}} \\ &= \sqrt{\frac{(17 - 1)\, 3.3^2 + (22 - 1)\, 2.8^2}{17 + 22 - 2}} \\ &= 3.0264 \end{aligned}$$

(b) Does the mean weight gain differ between the two groups? Carry out an appropriate hypothesis test. Be sure to give a range in which the P-value falls. Make sure you interpret the results of the test.

Answer:

Let μ_1 be the mean weight gain of growth hormone mice. Let μ_2 be the mean weight gain of control group mice. The appropriate hypothesis test is

$$H_0 : \mu_1 = \mu_2 \text{ vs } H_A : \mu_1 \neq \mu_2$$

The corresponding t statistic is

$$t = \frac{26.3 - 24.4}{3.0264\sqrt{\frac{1}{17} + \frac{1}{22}}} = 1.94$$

The P-value for this test will be given by

$$2P(T > 1.94)$$

since the df = 17 + 22 -2 = 37 which is not in the table to be conservative use df=30 line. We will put bounds on the P-value. First we note that

$$1.697 < 1.94 < 2.042$$

converting to P-values

$$0.05 > P(T > 1.94) > 0.025$$

and so

$$0.1 > 2P(T > 1.94) > 0.05$$

so we see that 0.05 < P-value < 0.1. We can not reject the null hypothesis. Therefore we would not be able to conclude that there is a difference in weight gain.

(c) Give the 95% confidence interval for the difference in mean weight gain between the two groups.

Answer:

The 95% confidence interval for $\mu_1 - \mu_2$ is given by

$$\bar{Y}_1 - \bar{Y}_2 \pm t^* \mathrm{SE}\left(\bar{Y}_1 - \bar{Y}_2\right)$$

and so

$$26.3 - 24.4 \pm 2.042\,(3.0264)\sqrt{\frac{1}{17} + \frac{1}{22}}$$

and so the 95% confidence interval is given by $(-0.096, 3.896)$

Question 2. (16 points)

A technician is interested in tuning a particular machine in a factory. He wants to tune the machine so that it is producing the smallest possible y values. To control the machine he can use a dial which sets the x value. He performs an experiment by setting the dial at different settings of x and then recording the y value. Consider the following data as the results of his experiment

y	x
2.80	-2
1.46	-1
1.99	0
8.75	1
17.47	2
3.25	-2
1.35	-1
2.73	0
7.20	1
16.40	2

(a) Are the functions 1, x, x^2 an orthogonal basis?

Answer:

$$\sum_{i=1}^{n} 1x_i = \sum_{i=1}^{n} x_i = -2-1+0+1+2-2-1+0+1+2 = 0$$

$$\sum_{i=1}^{n} 1x_i^2 = \sum_{i=1}^{n} x_i^2 = 4+1+0+1+4+4+1+0+1+4 = 20 \neq 0$$

$$\sum_{i=1}^{n} x_i x_i^2 = \sum_{i=1}^{n} x_i^3 = -8-1+0+1+8-8-1+0+1+8 = 0$$

Since 1 and x^2 are not orthogonal the basis is not orthogonal.

(b) Briefly describe how you would estimate the parameters of the regression model $\mu(x) = \beta_0 + \beta_1 x + \beta_2 x^2$. (you do not need to find $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$).

Answer:

We would use the method of least squares. Briefly this means we want to minimize the SSE ($\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - \hat{\beta}_2 x_i^2\right)^2$) with respect to $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$. We do this my differentiating the SSE and setting the resultant equations equal to 0. These equations are the normal equations. Solving the normal equations for $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ would give the parameter estimates.

(c) Consider the alternative basis 1, x, $x^2 - 2$. Is this an orthogonal basis?

Answer:

From part (a) we have shown that 1 and x are orthogonal.

$$\sum_{i=1}^{n} 1(x_i^2 - 2) = \sum_{i=1}^{n} (x_i^2 - 2) = 2-1-2-1+2+2-1-2-1+2 = 0$$

$$\sum_{i=1}^{n} x_i(x_i^2 - 2) = -4+1+0-1+4-4+1+0-1+4 = 0$$

so 1, x, $x^2 - 2$ is an orthogonal basis.

(d) Estimate the parameters of the alternative model $\tilde{\mu}(x) = \tilde{\beta}_0 + \tilde{\beta}_1 x + \tilde{\beta}_2 (x^2 - 2)$

Answer:

Since we have shown that we are using an orthogonal basis we can use the formula $\hat{\beta}_j = \frac{\sum_{i=1}^{n} g_j(\mathbf{x}_i) y_i}{\sum_{i=1}^{n} g_j^2(\mathbf{x}_i)}$ for $j = 0, \ldots, p$. And so

$$\hat{\tilde{\beta}}_0 = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} 1^2} = \bar{y} = 6.34$$

$$\hat{\tilde{\beta}}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} = \frac{68.78}{20} = 3.439$$

$$\hat{\tilde{\beta}}_2 = \frac{\sum_{i=1}^{n}\left(x_i^2-2\right)y_i}{\sum_{i=1}^{n}\left(x_i^2-2\right)^2} = \frac{51.64}{28} = 1.844$$

So the fitted model is $\hat{\mu}(x) = 6.34 + 3.439x + 1.844(x^2-2)$

(e) **Optional: Worth up to 5 bonus points** Explain how β_0, β_1 and β_2 relate to $\tilde{\beta}_0$, $\tilde{\beta}_1$ and $\tilde{\beta}_2$. Give an approximate answer for the tuning value x that the technician should choose.

Answer:

Both parameterizations should give the same fitted values so

$$\beta_0 + \beta_1 x + \beta_2 x^2 = \tilde{\beta}_0 + \tilde{\beta}_1 x + \tilde{\beta}_2(x^2-2)$$

equating the equivalent terms on each side gives us $\beta_0 = \tilde{\beta}_0 - 2\tilde{\beta}_2$, $\beta_1 = \tilde{\beta}_1$ and $\beta_2 = \tilde{\beta}_2$. Since the fitted model is

$$y = 6.34 + 3.439x + 1.844(x^2-2)$$

to minimize differentiate with respect to x gives

$$\frac{dy}{dx} = 3.439 + 2(1.844)x = 0$$

and so

$$x = -\frac{3.439}{2\,(1.844)} = -0.932$$