Math 326 – Homework 07 (10.10 – 11.5)
Due (via upload to Canvas) Friday, April 1, 2022 at 11:59 PM

1. Suppose that we have a random sample of n observations from the density function

$$f(y|\theta) = \frac{y^2 e^{-y/\theta}}{2\theta^3}$$

on support $y > 0$

(a) Determine the rejection region for the most powerful test of $H_0 : \theta = \theta_0$ vs $H_a : \theta = \theta_a$, assuming $\theta_a > \theta_0$.

Solution: We have $L(\theta) = \prod_1^n \frac{1}{2\theta^3} y_i^2 e^{-y_i/\theta} = \left(\frac{1}{2\theta^3}\right)^n \prod_1^n y_i^2 e^{-\sum_1^n y_i/\theta}$. Then

$$\frac{L(\theta_0)}{L(\theta_a)} = \frac{(\frac{1}{2\theta_0^3})^n \prod_1^n y_i^2 e^{-\sum_1^n y_i/\theta_0}}{(\frac{1}{2\theta_a^3})^n \prod_1^n y_i^2 e^{-\sum_1^n y_i/\theta_a}} = \left(\frac{\theta_a}{\theta_0}\right)^{3n} \exp\left[\sum_1^n y_i \left(\frac{1}{\theta_a} - \frac{1}{\theta_0}\right)\right] < k.$$

As $\theta_a > \theta_0$, the sign of $\frac{1}{\theta_a} - \frac{1}{\theta_0} < 0$. Thus, to make the ratio of likelihood functions small, we would need the statistic $T = \sum y_i$ to be large. That is, the rejection region will be defined by

$$T = \sum y_i > k'$$

for some k' chosen to satisfy a prescribed $1 - \alpha$ level.

Note that to use this in practice, we would need to determine how T is distributed. Since Y is distributed by Gamma$(3, \theta)$ (with the assumption that $\theta > 0$), by products of moment-generating functions ($m_T(t) = (1 - \theta t)^{-3n}$) we would get that T is distributed by Gamma$(3, \theta)$.

(b) Is the test you defined in part (a) uniformly most powerful for the alternative $\theta > \theta_0$? Briefly explain your answer.

Solution: Because the test found in part (a) does not depend on any specific $\theta_a > \theta_0$, the same argument will work of all $\theta_a > \theta_0$. Thus the test is a uniformly most powerful test.

The following table contains dietary data (calories and the content of fat, sodium, carbohydrate, and protein) in some standard hamburgers that can be found at local fast food restaurants.

	cal	fat (g)	sodium (mg)	carbs (g)	protein (g)
BK Jr.	310	18	390	27	13
Wendy's Jr.	250	11	420	25	13
McDonald's	250	9	480	31	12
Culvers	390	17	480	38	20
Steak-n-Shake	320	14	830	32	15
Sonic Jr.	330	16	610	32	15

2. We wish to explore if there is a relationship between fat and sodium. The conjecture is that leaner meat need more salt to enhance flavor.

   (a) Compute the least squares regression line with response variable sodium content and input variable fat content. Clearly state sums of the intermediate calculations: $\overline{x}, \overline{y}, S_{xy}, S_{xx}$.

Solution: I used Mathematica to compute these. (File is available upon request.) We see that $\overline{x} \approx 14.17$, $\overline{y} = 535$, $S_{xy} = 25$, $S_{xx} \approx 62.83$. Then $\hat{\beta}_1 = \dfrac{S_{xy}}{S_{xx}} \approx 0.398$ and $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} \approx 529.363$. The best-fit line is

$$\hat{y} = 529.363 + .398x.$$

   (b) Calculate S^2. Again, state any necessary intermediary sums.

Solution: $SSE \approx 132940.05$ and $S^2 = \dfrac{SSE}{6-2} \approx \dfrac{132940.05}{4} \approx 33,235$.

   (c) Calculate the correlation coefficient, ρ^2.

Solution: We have a couple of representations of ρ^2. Using $\rho^2 = \dfrac{S_{xy}^2}{S_{xx} \cdot S_{yy}}$ we get

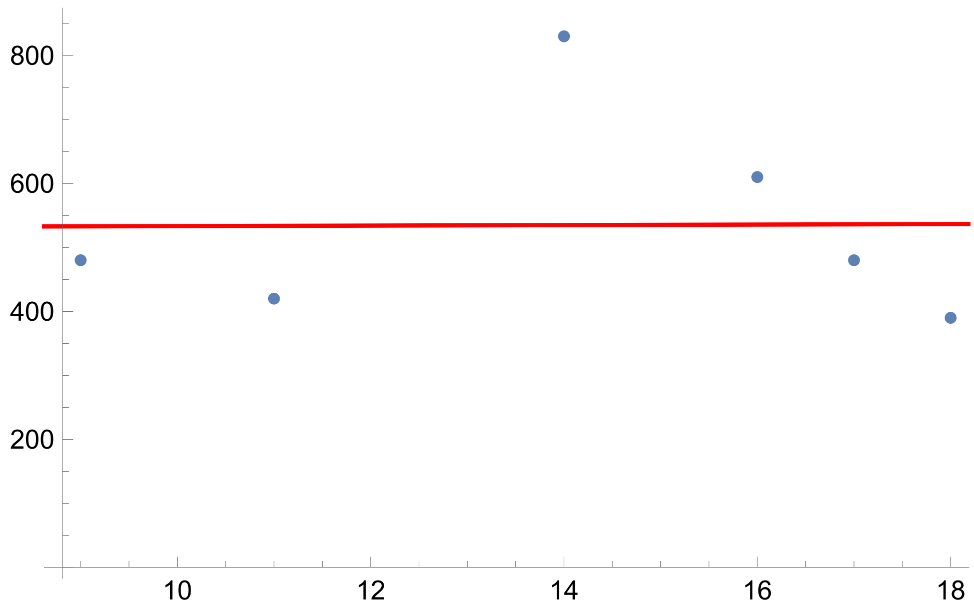
$$\rho^2 = \frac{25^2}{(377/6) \cdot 132950} \approx 0.0000749.$$

   (d) A good rule of thumb for the correlation coefficient in regards to best-fit lines is that if ρ^2 is greater than 0.70, then the line is a good model for the spread of the data. Is using a line a good model for this data?

Solution: $\rho^2 \approx .0000749 << 0.70$ and the line is unlikely to be a good model for this data. (As we clearly see in the next.)

(e) Sketch a scatterplot of the data and draw the best-fit line and interpret the picture in context of your answer in (d).

Solution:



Based on the graph and the size of ρ^2, we can conclude that fat and sodium do not have a linear relationship.

3. American culture is focused on fat intake as corresponding to a high-calorie diet.

(a) Compute the least squares regression line with response variable calorie count and input variable fat content. Clearly state sums of the intermediate calculations: $\overline{x}, \overline{y}, S_{xy}, S_{xx}$.

Solution: Using statistical software, we can see that $\overline{x} \approx 14.17$, $\overline{y} = 308.3$, $S_{xy} \approx 761.67$, $S_{xx} \approx 62.83$. Then $\hat{\beta}_1 = \dfrac{S_{xy}}{S_{xx}} \approx 12.122$ and $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} \approx 136.605$. Altogether, the best-fit line is

$$Y = 136.605 + 12.122x.$$

(b) Suppose a new burger on the market is known to have 650 calories. What is a good estimate for how much fat is in the burger?

Solution: According to our best fit line, $650 = 136.605 + 12.122x$. Solving for x, we see $x = 42.352$ grams of fat is a good estimate for a burger with 650 calories.

(c) Find an 80% confidence interval for the slope of the regression line.

Solution: An 80% confidence interval for $\hat{\beta}_1$ is $\hat{\beta}_1 \pm t_{0.10}(4) S \sqrt{\dfrac{1}{S_{xx}}}$ with 4 degrees of freedom. We can find using R that $S = \sqrt{\dfrac{SSE}{4}} = \sqrt{\dfrac{4850.3979}{4}} \approx 34.822$, $t_{0.10}(4) = 1.533$ for 4 degrees of freedom and $S_{xx} \approx 62.83$. This yields the 80% confidence interval for $\hat{\beta}_1$

$$12.122 \pm 1.533(34.822)(0.126) = 12.122 \pm 6.735 \text{ or } (5.387, 18.857).$$

(d) Is there statistical evidence that the slope of the regression line is greater than 10? Run a hypotheses test at $\alpha = 0.05$.

Solution: Our hypothesis test is:

$$H_0 : \beta_1 = 10 \text{ vs } H_a : \beta_1 > 10.$$

We will then look at $P(\hat{\beta}_1 > 12 | \beta = 10)$. This is equivalent to $P(T > \dfrac{\hat{\beta}_1 - \beta}{S\sqrt{\dfrac{1}{S_{xx}}}}) = P(T > \dfrac{2}{4.388}) = P(T > .456) \approx .336$. Since $.336 > .05$ there is not sufficient evidence that the slope of the regression line is greater than 10.