MTH375: Mathematical Statistics - Homework #10

Cason Konzer

April 11, 2022

Key Concepts: χ^2 test for goodness of fit, linear models, estimators of β, σ^2 and their distributions.

1. A six-sided die is rolled 100 times. The numbers that come up are ...

number	1	2	3	4	5	6
frequency	20	15	17	11	18	19

We want to use the χ^2 test to decide whether there is suffcient evidence at level of significance $\alpha = 0.05$ to conclude that the die is unfair, i.e., that some numbers are more likely than others.

State H_0 and H_A in statistical language, describe the test you will use, carry out the test, and state the conclusion.

Solution:

We want to test for a (un)fair die, and thus that all sides have equal probablility, are hypotheses are thus ...

$$H_0 : p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = 1/6 \text{ vs. } H_A : \neg H_0.$$

Our test is know as a χ^2 test for categorical data, as we have 6 categories we test on a χ^2 statistic with 5 degrees of freedom.

Now to carry out the test in **R** ...

- `c(20, 15, 17, 11, 18, 19) -> Oi` .
- `c(100, 100, 100, 100, 100, 100) / 6 -> Ei` .
- `sum( (Oi - Ei)^2 / Ei ) -> T` .
- `qchisq(p = (1 - 0.05), df = 5) -> t` .

Our test is of the form "Reject H_0 if $T > t$".

Our results are $T = 3.2$ and $t = 11.0705$
As $3.2 < 11.0705$ we fail to reject H_0 with a 95% confidence.

2. Use the χ^2 test to decide, at level of significance $\alpha = 0.05$, whether the data supports or refutes the claim that the 60 data points below came from an exponentially distributed population.

$$\begin{gathered}0.105,\ 0.183,\ 0.219,\ 0.313,\ 0.326,\ 0.345,\ 0.454,\ 0.461,\ 0.467,\ 0.551, \\ 0.603,\ 0.757,\ 0.802,\ 0.824,\ 0.826,\ 0.844,\ 0.987,\ 1.087,\ 1.159,\ 1.180, \\ 1.249,\ 1.252,\ 1.317,\ 1.326,\ 1.390,\ 1.398,\ 1.580,\ 1.618,\ 1.653,\ 1.660, \\ 1.759,\ 1.850,\ 1.875,\ 2.638,\ 2.691,\ 2.811,\ 2.823,\ 2.828,\ 2.924,\ 3.108, \\ 3.323,\ 3.671,\ 3.792,\ 3.797,\ 4.574,\ 4.855,\ 4.924,\ 5.098,\ 5.287,\ 5.346, \\ 6.000,\ 6.335,\ 6.491,\ 6.625,\ 7.125,\ 7.586,\ 8.028,\ 9.071,\ 10.783,\ 11.034\end{gathered}$$

The data was put in order from least to greatest and sorted into 6 groups of 10 only to make it easier to read. The sum of these numbers is 176.0.

Divide the positive real numbers into 5 intervals in such a way that the probability of landing in a given interval is 0.2, and use those intervals to perform the test.

Since you will be estimating a parameter, you will substract one additional degree to freedom when performing the χ^2 test.

Solution:

We first need the MLE for a random variable distibuted on $Exponential(\lambda)$.

This is $\hat{\lambda}_{MLE} = \overline{X} = \Sigma_X / n = 176/60$.

We thus are testing the following hypotheses ...

$$H_0 : X_i \sim Exponential(\lambda = 176/60) \text{ vs. } H_A : \neg H_0.$$

Our test is know as a χ^2 test for categorical data, as we have 5 categories we test on a χ^2 statistic with 4 degrees of freedom.

We can now use **R** to compute our data categories of equal probability interval 0.2, and then carry out the test, leveraging the expected outcomes from our interval assumption ...

- `pexp(q = c(0.2, 0.4, 0.6, 0.8, 1.0), rate = 60 / 176) * 60 -> cuts` .

We find the following cuts, $(0, 3.955), (3.955, 7.648), (7.648, 11.099), (11.099, 14.322), (14.322, 17.333)$, from which we can then count our observations for each category.

- `c(44, 12, 4, 0, 0) -> Oi` .
- `c(60, 60, 60, 60, 60) / 5 -> Ei`.
- `sum( (Oi - Ei)^2 / Ei ) -> T` .
- `qchisq(p = (1 - 0.05), df = 4) -> t` .

Our test is of the form "Reject H_0 if $T > t$".

Our results are $T = 114.6667$ and $t = 9.487729$
As $114.666 > 9.487729$ we reject H_0 with a 95% confidence.

3. Find the equation of the line that best fits the points ...

$$\{(-4, 5), (-4, 4), (-3, 2), (-2, 3), (-2, 1), (0, 1), (0, 2), (1, 3), (1, 0), (2, 0), (2, -1), (3, 1)\}$$

(in the sense of least squares). Sketch a graph containing these points and your line.

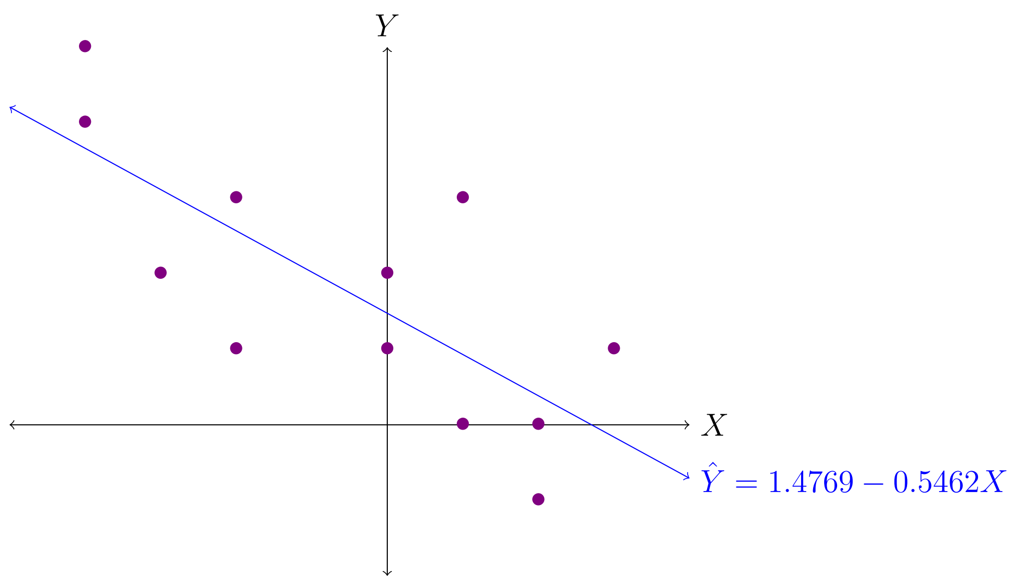
Solution:

We can use **R** to compute our regression.

- `c(-4, -4, -3, -2, -2, 0, 0, 1, 1, 2, 2, 3) -> X` .
- `c(5, 4, 2, 3, 1, 1, 2, 3, 0, 0, -1, 1) -> Y` .
- `lm(Y~X) -> reg` .
- `plot(reg)` .

R outputs our coefficients $\hat{\beta}_0 = 1.4769$ and $\hat{\beta}_1 = -0.5462$, thus our equation is $\hat{Y} = 1.4769 - 0.5462X$.

Sketching the situation ...



4. Suppose the data in #3 came from a population satisfying the model in which $Y_1, \ldots, Y_{12}$ are independent, $\sim Normal(\mu_k = \beta_0 + \beta_1 x_k, \sigma^2)$. Test the hypotheses at level of significance $\alpha = 0.01 \ldots$

$$H_0 : \beta_1 = 0 \text{ vs. } H_A : \beta_1 \neq 0.$$

Solution:

From the previous problem we can conduct this test easily in **R** with the command `summary(reg)` .

From the command we can see that the p-value given when conducting a two tailed t-test against the null hypothesis is 0.00213. Thus the hypothesis is rejected for $\alpha > 0.00123$, hence we reject the H_0 at $\alpha = 0.01$.