Review of Introduction to Probability and Statistics

Chris Mack, http://www.lithoguru.com/scientist/statistics/review.html

Practice Final Exam Solutions

This exam is closed books and notes. You can use the equation summary sheet (3 pages) and probability distribution tables found on the webpage. Calculators allowed (and encouraged).

1. Please calculate summary statistics for an oil well's daily production, in barrels.

   214, 203, 226, 198, 243, 225, 207, 203, 208

   a. Mean
   b. Median
   c. Standard Deviation
   d. 90% confidence interval around the mean

Solution:

Mean =	214.11 barrels
Median =	208.0 barrels
Standard Deviation =	14.5 barrels
90% critical t-score =	1.86
Margin of Error =	9.0 barrels
CI Low =	205.1 barrels
CI High =	223.1 barrels

2. Please calculate least-squares linear regression statistics (slope, intercept, goodness of fit) for the pressure (x) vs. flow rate (y) data below. Summary statistics have been precalculated for your convenience.

Pressure	5	6	7	8	9	10
Flow	14	25	70	85	49	105

$\bar{x} = 7.5, \quad s_x = 1.871, \quad \bar{y} = 58, \quad s_x = 35.18, \quad cov(x,y) = 54.2$

Solution:

$$a = \frac{cov(X,Y)}{var(X)} = \frac{54.2}{(1.871)^2} = 15.5$$

$$b = \bar{y} - a\bar{x} = 58 - 15.5(7.5) = -58.1$$

$$R^2 = \frac{[cov(X,Y)]^2}{var(X)var(Y)} = \left(\frac{54.2}{(1.871)(35.18)}\right)^2 = 0.68$$

3. For the least-squares best fit line of problem 2, the standard deviation of the residuals was found to be 22.31. Use this and other information from problem 2 to calculate the 95% confidence intervals for your estimates of the slope and intercept.

Solution:

Since n = 6, we must use a Student's t distribution. For 4 degrees of freedom and $\alpha = 0.05$, the critical t-value is 2.776 (from the Student's t Table, two-sided distribution)

$$var(a) = \frac{var(\epsilon)}{n\, var(X)}, \qquad SE(a) = \frac{22.31}{\sqrt{6}(1.871)} = 4.87, \qquad ME(a) = 2.776(4.87) = 13.5$$

Slope 95% CI: (2.0, 29.0)

$$var(b) = \frac{var(\epsilon)}{n}\left(1 + \frac{\bar{x}^2}{var(X)}\right), \qquad SE(b) = \frac{22.31}{\sqrt{6}}\sqrt{1 + \left(\frac{7.5}{1.871}\right)^2} = 37.6$$

$$ME(b) = 2.776(37.6) = 105$$

Intercept 95% CI: (-162, 46)

4. For the least-squares best fit line of problems 2 and 3, calculate the predicted value for x = 8. Show the 95% confidence interval for this prediction.

Solution:

$$\hat{y}_i = ax_i + b = 15.5(8) - 58.1 = 65.7$$

$$var(\hat{y}_i) = \frac{var(\epsilon)}{n}\left(1 + \frac{(x_i - \bar{x})^2}{var(X)}\right), \qquad SE(\hat{y}_i) = \frac{22.31}{\sqrt{6}}\sqrt{1 + \left(\frac{8 - 7.5}{1.871}\right)^2} = 9.43$$

$$ME(\hat{y}_i) = 2.776(9.43) = 26.2$$

Predicted value 95% CI: (39.6, 91.9)

5. A bowl contains 10 marbles, four of which are blue. What is the probability that two marbles chosen at random will both be blue?

Solution:

Let A_1 = event that the first ball is blue, and A_2 = event that second ball is blue.

$\mathbb{P}(A_1) = 0.4$, $\mathbb{P}(A_2|A_1) = 3/9$, $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\,\mathbb{P}(A_2|A_1) = 4/30 = 0.133$.

6. Consider a family photo where the grandmother is to be in the center of a line of people. Given 7 family members in the photo (including grandmother), how many different ways are there to line up the people?

Solution:

Consider two stages: arrange everyone except the grandmother, then insert the grandmother. For the first stage, there are 6! Permutations. For the second stage, there is only one way to put the grandmother in the middle. Thus, the total number of permutations is 6! = 720.

7. I have 20 students in my class, and I randomly assign two of them to clean my erasers after class. How many different pairs of eraser cleaners are there?

Solution:

How many ways can I choose 2 from 20?

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}, \qquad \binom{20}{2} = \frac{20!}{2!\,(20-2)!} = \frac{20 \cdot 19}{2} = 190$$

8. The number of bad checks that a bank receives each day can be modelled as a Poisson distribution with λ = average number of bad checks received per day. If $\lambda = 6$, what is the probability that 4 bad check will be received on any given day?

Solution:

$$p_X(k) = \frac{\lambda^k}{k!}e^{-\lambda}, \qquad p_X(4) = \frac{6^4}{4!}e^{-6} = 0.134$$

9. The humidity in a controlled environment is modeled as a normal random variable with mean and standard deviation equal to 40% and 2.2%, respectively. What is the probability that the humidity will exceed 45%?

Solution:

$$Z = \frac{X - \mu}{\sigma} \sim N(0,1), \qquad z = \frac{45 - 40}{2.2} = 2.273$$

From a standard normal table, $\mathbb{P}(z > 2.273) = 1 - \mathbb{P}(z < 2.273) = 1 - 0.9885 = 0.0115$, or about a 1% chance.

10. Two neighborhoods disagree about whether people in the first neighborhood are more energy conscious than the second. Below are the statistics for last June's electricity usage (in kW-h) at the houses of these two neighborhoods. Perform an hypothesis test to help settle this dispute.

	Mean	Std. Dev.	Number of houses
Neighborhood 1	1216	188	55
Neighborhood 2	1307	170	42

Solution:

H_0: $\mu_2 - \mu_1 = 0$ H_A: $\mu_2 - \mu_1 > 0$

This is a one-tailed test, and the sample sizes are large enough that the sampling distributions will be about normal.

$$z = \frac{\bar{x}_2 - \bar{x}_1 - 0}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}} = \frac{1307 - 1216}{\sqrt{\frac{188^2}{55} + \frac{170^2}{42}}} = \frac{91}{36.5} = 2.49$$

The p-value for this one-tailed test is 0.006. Thus, unless we picked a very small significance level, we can reject the null hypothesis in favor of the alternative hypothesis. The people in Neighborhood 1 really do use less electricity than the residents of neighborhood 2, with 99.4% confidence.