

## 1 Iterated Expectations

1. Given  $E[X|Y] = 2Y$  and  $f(Y) = 0.5$  with  $Y \in [-3, -1]$ , what is  $E[X]$ ?

**Solution:** We use the law of iterated expectations:

$$E[X] = E[E[X|Y]] = E[2Y] = 2E[Y]$$

Since  $f(Y) = 0.5$ ,  $Y$  is uniformly distributed on  $[-3, -1]$ . The expectation of  $Y$  is:

$$E[Y] = \frac{-3 + (-1)}{2} = -2$$

Thus:

$$E[X] = 2(-2) = -4$$

2. Given  $E[Z|H] = 15H - 10$  and  $H \sim \text{Bernoulli}(0.2)$ , what is  $E[Z]$ ?

**Solution:** Again, by the law of iterated expectations:

$$E[Z] = E[E[Z|H]] = E[15H - 10] = 15E[H] - 10$$

Since  $H \sim \text{Bernoulli}(0.2)$ , we have  $E[H] = 0.2$ . Therefore:

$$E[Z] = 15(0.2) - 10 = 3 - 10 = -7$$

## 2 Properties of Estimators

1. Let  $X_1, \dots, X_n \sim \text{Poisson}(\lambda)$  and let  $\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n X_i$ . Find the bias, standard error, and MSE of this estimator.

**Solution:** - The expectation of  $\hat{\lambda}$  is:

$$E[\hat{\lambda}] = E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \lambda$$

Thus,  $\hat{\lambda}$  is an unbiased estimator. The variance of  $\hat{\lambda}$  is:

$$\text{Var}(\hat{\lambda}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{\lambda}{n}$$

The standard error (SE) is:

$$SE(\hat{\lambda}) = \sqrt{\frac{\lambda}{n}}$$

The mean squared error (MSE) is:

$$\text{MSE}(\hat{\lambda}) = \text{Var}(\hat{\lambda}) + \text{Bias}(\hat{\lambda})^2 = \frac{\lambda}{n}.$$

2. Let  $X_1, \dots, X_n \sim \text{Uniform}(0, \theta)$  and let  $\hat{\theta} = 2\bar{X}_n$ . Find the bias, standard error, and MSE of this estimator.

**Solution:** - The expectation of  $\hat{\theta}$  is:

$$E[\hat{\theta}] = E[2\bar{X}_n] = 2E[\bar{X}_n] = 2\left(\frac{\theta}{2}\right) = \theta.$$

Thus,  $\hat{\theta}$  is an unbiased estimator. The variance of  $\hat{\theta}$  is:

$$\text{Var}(\hat{\theta}) = 4\text{Var}(\bar{X}_n) = 4\left(\frac{\theta^2}{12n}\right) = \frac{\theta^2}{3n}$$

The standard error (SE) is:

$$SE(\hat{\theta}) = \sqrt{\frac{\theta^2}{3n}} = \frac{\theta}{\sqrt{3n}}$$

The mean squared error (MSE) is:

$$\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + \text{Bias}(\hat{\theta})^2 = \frac{\theta^2}{3n}.$$

### 3 Birds of a Feather Get Their News on Twitter (X)

A poll conducted in 2013 found that 52% of U.S. adult Twitter users get at least some news on Twitter. The standard error for this estimate was 2.4%, and a normal distribution may be used to model the sample proportion.

1. Construct a 99% confidence interval for the fraction of U.S. adult Twitter users who get some news on Twitter, and interpret the confidence interval in context.

**Solution:** To construct a 99% confidence interval for the proportion, we use the formula:

$$CI = \hat{p} \pm z_{\alpha/2} \cdot SE$$

where  $\hat{p} = 0.52$ ,  $SE = 0.024$ , and  $z_{\alpha/2}$  is the critical value for a 99% confidence level. From standard normal tables,  $z_{\alpha/2} = 2.576$ .

Thus, the confidence interval is:

$$CI = 0.52 \pm 2.576 \cdot 0.024 = 0.52 \pm 0.06182 = [0.45818, 0.58182]$$

Interpretation: We are 99% confident that the true proportion of U.S. adult Twitter users who get at least some news from Twitter is between 45.82% and 58.18%.

2. (T/F) The data provide statistically significant evidence that more than half of U.S. adult Twitter users get some news through Twitter. Use a significance level of  $\alpha = 0.01$ .

**Solution:** The null hypothesis is  $H_0 : p = 0.5$  and the alternative hypothesis is  $H_A : p > 0.5$ . We calculate the test statistic:

$$Z = \frac{\hat{p} - 0.5}{SE} = \frac{0.52 - 0.5}{0.024} = \frac{0.02}{0.024} = 0.8333$$

The corresponding p-value is:

$$P(Z > 0.8333) = 1 - P(Z \leq 0.8333) = 1 - 0.7967 = 0.2033$$

Since the p-value is greater than 0.01, we fail to reject the null hypothesis. Therefore, the data do not provide statistically significant evidence that more than half of U.S. adult Twitter users get some news through Twitter.

3. (T/F) Since the standard error is 2.4%, we can conclude that 97.6% of all U.S. adult Twitter users were included in the study.

**Solution:** False. The standard error is a measure of the variability of the sample proportion, not the percentage of people included in the study.

4. (T/F) If we want to reduce the standard error of the estimate, we should collect less data.

**Solution:** False. Collecting more data reduces the standard error, not less.

5. (T/F) If we construct a 90% confidence interval for the percentage of U.S. adult Twitter users who get some news through Twitter, this confidence interval will be wider than a corresponding 99% confidence interval.

**Solution:** False. A 90% confidence interval will be narrower than a 99% confidence interval because the critical value for a 90% confidence level is smaller than that for a 99% confidence level.

## 4 Choose Your Own Death

**Problem 5:** There is a theory that people can postpone their death until after an important event. To test the theory, Phillips and King (1988) collected data on deaths around the Jewish holiday Passover. Of 1919 deaths, 922 died the week before the holiday and 997 died the week after. Think of this as a binomial and test the null hypothesis that  $\theta = \frac{1}{2}$ . Report and interpret the p-value. Also construct a confidence interval for  $\theta$ .

**Solution:**

**Hypothesis Test:** We are testing the null hypothesis:  $H_0 : \theta = \frac{1}{2}$  against the alternative hypothesis:  $H_A : \theta \neq \frac{1}{2}$ . The test statistic for a binomial test is:

$$Z = \frac{\hat{\theta} - \frac{1}{2}}{\sqrt{\frac{1}{4n}}}$$

where  $\hat{\theta} = \frac{997}{1919} = 0.5195$  and  $n = 1919$ . First, calculate the standard error:

$$SE = \sqrt{\frac{1}{4n}} = \sqrt{\frac{1}{4 \times 1919}} = \sqrt{\frac{1}{7676}} \approx 0.0114.$$

Now, calculate the Z-score:

$$Z = \frac{0.5195 - 0.5}{0.0114} = \frac{0.0195}{0.0114} \approx 1.71.$$

The p-value corresponding to  $Z = 1.71$  for a two-tailed test is:

$$p = 2 \times P(Z > 1.71) = 2 \times (1 - 0.9564) = 2 \times 0.0436 = 0.0872$$

Since the p-value 0.0872 is greater than any reasonable significance level (e.g.,  $\alpha = 0.05$ ), we fail to reject the null hypothesis. There is not enough evidence to support the theory that people can postpone their death until after an important event.

**Confidence Interval:**

A 95% confidence interval for  $\theta$  can be constructed using the formula:

$$CI = \hat{\theta} \pm z_{\alpha/2} \cdot SE$$

Using  $z_{\alpha/2} = 1.96$  for a 95% confidence level:

$$CI = 0.5195 \pm 1.96 \cdot 0.0114 = 0.5195 \pm 0.0224 = [0.4971, 0.5419]$$

**Interpretation:** We are 95% confident that the true proportion of people who die the week after Passover is between 49.71% and 54.19%. Since 50% is within this interval, we do not have enough evidence to reject the null hypothesis that  $\theta = \frac{1}{2}$ .

## 5 Dating on College Campuses

A survey conducted on a reasonably random sample of 203 undergraduates asked about the number of exclusive relationships these students have been in. The histogram shows the distribution of the data from this sample. The sample average is 3.2 with a standard deviation of 1.97.

Estimate the average number of exclusive relationships undergraduate students have been in using the Normal distribution and a 90% confidence interval, and interpret this interval in context.

**Solution:** We can estimate the population mean using a confidence interval for the sample mean. The formula for a confidence interval for the mean is:

$$CI = \bar{X} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

where  $\bar{X} = 3.2$ ,  $s = 1.97$ , and  $n = 203$ . For a 90% confidence level, the critical value  $z_{\alpha/2} = 1.645$ .

First, calculate the standard error:

$$SE = \frac{s}{\sqrt{n}} = \frac{1.97}{\sqrt{203}} \approx \frac{1.97}{14.24} \approx 0.1384$$

Now, construct the confidence interval:

$$CI = 3.2 \pm 1.645 \times 0.1384 = 3.2 \pm 0.2277 = [2.9723, 3.4277]$$

**Interpretation:** We are 90% confident that the true average number of exclusive relationships undergraduate students have been in is between 2.97 and 3.43.

## 6 Statistical Significance

Determine whether the following statement is true or false, and explain your reasoning: "With large sample sizes, even small differences between the null value and the point estimate can be statistically significant."

**Solution:**

True. As sample size increases, the standard error of the estimate decreases, which makes the test statistic (e.g., Z or t) larger for a given difference between the point estimate and the null value. As a result, even small differences between the point estimate and the null hypothesis can lead to a statistically significant result because the p-value will become smaller. This phenomenon occurs because the confidence in the precision of the estimate increases with a larger sample size.

## 7 Sleep Deprivation

New York is known as "the city that never sleeps." A random sample of 25 New Yorkers were asked how much sleep they get per night. Statistical summaries of these data are shown below. Do these data provide strong evidence that New Yorkers sleep less than 8 hours a night on average?

$$n = 25, \quad \bar{X} = 7.73, \quad s = 0.77, \quad \min = 6.17, \quad \max = 9.78$$

1. Write the hypotheses in symbols and in words.

**Solution:** We are testing whether New Yorkers sleep less than 8 hours per night on average. Null hypothesis:  $H_0 : \mu = 8$  (The average sleep duration is 8 hours per night.). Alternative hypothesis:  $H_A : \mu < 8$  (The average sleep duration is less than 8 hours per night.)

2. Calculate the test statistic,  $T$ , and the associated degrees of freedom.

**Solution:** The test statistic for a one-sample t-test is calculated as:

$$T = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}}$$

where  $\bar{X} = 7.73$ ,  $\mu_0 = 8$ ,  $s = 0.77$ , and  $n = 25$ . First, calculate the standard error:

$$SE = \frac{s}{\sqrt{n}} = \frac{0.77}{\sqrt{25}} = \frac{0.77}{5} = 0.154$$

Now, calculate the test statistic:

$$T = \frac{7.73 - 8}{0.154} = \frac{-0.27}{0.154} \approx -1.7532$$

The degrees of freedom (df) is:

$$df = n - 1 = 25 - 1 = 24.$$

- Find and interpret the p-value in this context.

**Solution:** To find the p-value, we use a t-distribution with 24 degrees of freedom. The test is one-tailed (since we are testing  $H_A : \mu < 8$ ).

Using a t-table or calculator, the p-value for  $T = -1.7532$  and 24 degrees of freedom is approximately:

$$p \approx 0.045$$

Interpretation: The p-value is 0.045, which indicates that there is a 4.5% chance of observing a sample mean as extreme as 7.73 hours (or lower) if the true average sleep time were 8 hours. Since this p-value is less than 0.05, we have moderate evidence against the null hypothesis.

- What is the conclusion of the hypothesis test?

**Solution:** Since the p-value  $\approx 0.045$  is less than the significance level  $\alpha = 0.05$ , we reject the null hypothesis. There is moderate evidence to suggest that New Yorkers sleep less than 8 hours per night on average.

- If you were to construct a 90% confidence interval that corresponded to this hypothesis test, would you expect 8 hours to be in the interval?

**Solution:** Since the p-value is less than 0.05 and we rejected the null hypothesis, we would not expect 8 hours to be within a 90% confidence interval for the mean sleep time. The rejection of the null hypothesis suggests that the true mean is likely less than 8 hours.

## 8 Interpreting Public Opinion Polls

On June 28, 2012, the U.S. Supreme Court upheld the much-debated 2010 healthcare law, declaring it constitutional. A Gallup poll released the day after this decision indicates that 46% of 1,012 Americans agree with this decision. At a 95% confidence level, this sample has a 3% margin of error. Based on this information, determine if the following statements are true or false, and explain your reasoning.

- We are 95% confident that between 43% and 49% of Americans in this sample support the decision of the U.S. Supreme Court on the 2010 healthcare law.

**Solution:** False. The margin of error applies to the entire population of Americans, not just the sample. We are 95% confident that the true proportion of all Americans who support the decision lies between 43% and 49%, not just within the sample.

- We are 95% confident that between 43% and 49% of Americans support the decision of the U.S. Supreme Court on the 2010 healthcare law.

**Solution:** True. This statement correctly interprets the margin of error and confidence level. We are 95% confident that the true proportion of all Americans who support the decision is between 43% and 49%.

- If we considered many random samples of 1,012 Americans, and we calculated the sample proportions of those who support the decision of the U.S. Supreme Court, 95% of those sample proportions will be between 43% and 49%.

**Solution:** True. This statement correctly describes the interpretation of a 95% confidence interval. In repeated sampling, 95% of the calculated confidence intervals will contain the true proportion.

4. The margin of error at a 90% confidence level would be higher than 3%.

**Solution:** False. The margin of error decreases as the confidence level decreases. A 90% confidence interval would have a smaller margin of error compared to a 95% confidence interval.

mm