MATH 205: Statistical methods

November 8th, 2021

Lecture 18: Significance of Evidence

Chapter 7: Significance of evidence

Last week

Example

A sample of 66 obese adults was put on a low-carbohydrate diet for a year. The (sample) average weight loss was 7.7 lb and the sample standard deviation was 19.1 lb. Calculate a 99% lower confidence bound for the true average weight loss.

Question: How confident are you in the statement that the diet is effective?

Last week

- Idea: Construct the largest confidence interval I_{max} that doesn't contain 0

$$0 \notin (\bar{x} - z_\alpha \frac{s}{\sqrt{n}}, \infty)$$

- That corresponds to

$$\bar{x} = z_\alpha \frac{s}{\sqrt{n}}$$

or

$$z_\alpha = \frac{\bar{x}}{s/\sqrt{n}}$$

- (Informally) $1 - \alpha$ will be referred to as the $p - value$

Hypothesis testing

In a hypothesis-testing problem, there are two contradictory hypotheses under consideration

- The null hypothesis, denoted by H_0, is the claim that is initially assumed to be true
- The alternative hypothesis, denoted by H_a, is the assertion that is contradictory to H_0.
- The null hypothesis will be rejected in favor of the alternative hypothesis only if sample evidence suggests that H_0 is false.
- If the sample does not strongly contradict H_0, we will continue to believe in the probability of the null hypothesis.

Hypothesis testing: an analogy

In a criminal trial, there are to contradictory assertions

- the accused individual is innocent
- the accused individual is guilty

→ the claim of innocence is the favored or protected hypothesis

Hypothesis testing: example

- suppose a company is considering putting a new additive in the dried fruit that it produces
- the true average shelf life with the current additive is known to be 200 days
- With μ denoting the true average life for the new additive, the company would not want to make a change unless evidence strongly suggested that μ exceeds 200
- Null hypothesis:

$$H_0 : \mu = 200$$

- Alternative hypothesis:

$$H_a : \mu > 200$$

Test about a population mean

- Null hypothesis
$$H_0 : \mu = \mu_0$$
- The alternative hypothesis will be either:
  - $H_a : \mu > \mu_0$
  - $H_a : \mu < \mu_0$
  - $H_a : \mu \neq \mu_0$

Example 1

Problem

The drying time of a certain type of paint under specified test conditions is known to be normally distributed with mean value 75 min and standard deviation 9 min. Chemists have proposed a new additive designed to decrease average drying time. It is believed that drying times with this additive will remain normally distributed with $\sigma = 9$*. Because of the expense associated with the additive, evidence should strongly suggest an improvement in average drying time before such a conclusion is adopted.*

Test about a population mean

- Imagine we hypothesize that the average human body temperature is 94F.
- We collect temperature measurements from a random sample of n people.
- The mean of this sample is unlikely to be 94F.
- We must now find what caused the difference between the sample mean and the value we hypothesized.

Test about a population mean

- Two possibilities:
  - The hypothesis might be wrong.
  - It could be right, and the difference might just be because the sample is randomly chosen.
- One strategy is to construct an interval around the sample mean within which the true value will lie for (say) 99% of possible samples.
- If 94 is outside that interval, then:
  - the sample we obtained is inconsistent with the idea that the average human body temperature is 94F
  - if you want to believe the average human body temperature is 94F, you have to believe that you obtained a very odd sample.

"Proof by contradiction"

Ideas:

- Assume that a statement is true
- we then prove that this leads to a contradiction
- thus, we deduce that the original statement is wrong

Statistical "Proof by contradiction"

Ideas:

- Assume that a hypothesis (the null hypothesis) is true
- We ask ourselves, what is the probability that we'll see a dataset as contradictory as (or more contradictory than) the current one?
- That probability is referred to as the p-value (also called **observed significance level**) of the test
- If the p-value is less than a predetermined threshold (called **significant level**, often denoted by α), then we reject the null hypothesis

Note: "contradictory" is a relative concept and is reflected through the alternative hypothesis

z-test

- Given a random sample of size n from a distribution with mean μ (unknown) and either
  - the distribution is normal and population standard deviation σ is known, or
  - sample size $n > 40$
- Assume that the data is collected with the measured sample mean $\bar{x}$ and we want to test

$$H_0 : \mu = \mu_0$$
$$H_a : \mu < \mu_0$$

z-test: normal distribution with known σ

- If the null hypothesis $\mu = \mu_0$ is true, then $X_i \sim N(\mu_0, \sigma^2)$
- A sample would be more contradictory to the null hypothesis than the current sample we have if

$$\bar{X} \leq \bar{x} \quad \text{or} \quad \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \leq \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

- Thus, the p-value in this case is

$$P\left[Z \leq \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right] = \Phi\left(\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right)$$

P-values for z-tests

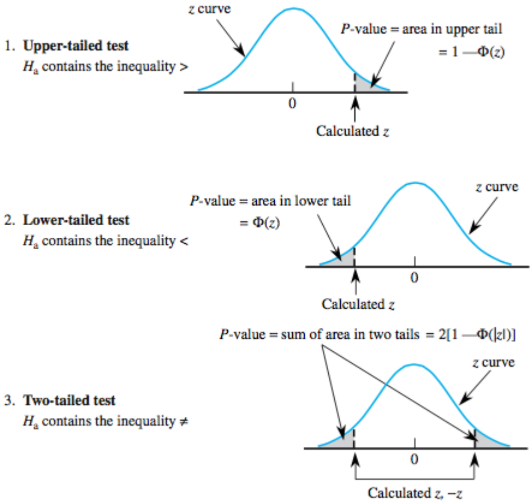

Figure 9.7 Determination of the P-value for a z test

Practice problem

Problem

The target thickness for silicon wafers used in a certain type of integrated circuit is 245 μm. A sample of 50 wafers is obtained and the thickness of each one is determined, resulting in a sample mean thickness of 246.18 μm and a sample standard deviation of 3.60 μm.

At signififcant level $\alpha = 0.01$, does this data suggest that true average wafer thickness is something other than the target value?

$\Phi(z)$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9278	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997

P-values for z-tests

1. Parameter of interest: μ = true average wafer thickness
2. Null hypothesis: H_0: $\mu = 245$
3. Alternative hypothesis: H_a: $\mu \neq 245$
4. Formula for test statistic value: $z = \dfrac{\bar{x} - 245}{s/\sqrt{n}}$
5. Calculation of test statistic value: $z = \dfrac{246.18 - 245}{3.60/\sqrt{50}} = 2.32$
6. Determination of P-value: Because the test is two-tailed,

$$P\text{-value} = 2[1 - \Phi(2.32)] = .0204$$

7. Conclusion: Using a significance level of .01, H_0 would not be rejected since .0204 > .01. At this significance level, there is insufficient evidence to conclude that true average thickness differs from the target value.

P-values for z-tests

$$P\text{-value:} \quad P = \begin{cases} 1 - \Phi(z) & \text{for an upper-tailed test} \\ \Phi(z) & \text{for a lower-tailed test} \\ 2[1 - \Phi(|z|)] & \text{for a two-tailed test} \end{cases}$$

Practice problem

Problem

The target thickness for silicon wafers used in a certain type of integrated circuit is 245 μm. A sample of 50 wafers is obtained and the thickness of each one is determined, resulting in a sample mean thickness of 246.18 μm and a sample standard deviation of 3.60 μm.

Does this data suggest that true average wafer thickness is larger than the target value? Carry out a test of significance at level .01 and provide the corresponding P-value.

Interpreting P-values

A P-value:

- is not the probability that H_0 is true
- is not the probability of rejecting H_0
- is the probability, calculated assuming that H_0 is true, of obtaining a test statistic value at least as contradictory to the null hypothesis as the value that actually resulted

Problem 1

A company that makes cola drinks states that the mean caffeine content per one 12-ounce bottle of cola is 40 milligrams. You work as a quality control manager and are asked to test this claim. During your tests, you find that a random sample of 30 bottles of cola (12-ounce) has a mean caffeine content of 39.2 milligrams. From a previous study, you know that the standard deviation of the population is $\sigma = 7.5$ milligrams. We assume that the caffeine content is normally distributed.

(a) (20 points) At α= 1% level of significant, can you reject the company's claim? What is the P-value associated with the test?