STAT 511

Lecture 15: Tests about Population Means and Population Proportions

Devore: Section 8.2-8.3

Prof. Michael Levine

March 28, 2020

A Normal Population with known σ

- This case is not common in practice. We will use it to illustrate basic principles of test procedure design
- Let $X_1, \ldots, X_n$ be a sample size n from the normal population. The null value of the mean is usually denoted μ_0 and we consider testing either of the three possible alternatives $\mu > \mu_0$, $\mu < \mu_0$ and $\mu \neq \mu_0$
- The test statistic that we will use is

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

It measures the distance of $\bar{X}$ from μ_0 in standard deviation units.

- Consider $H_a : \mu > \mu_0$ as an alternative. This is called an *upper-tailed* alternative.
- The outcome that would allow us to reject the null hypothesis $H_0 : \mu = \mu_0$ is that z is sufficiently larger than μ_0
- We start with calculating the P-value

$$P - \text{ value } = P(Z \geq z | H_0)$$

- Let us assume that $z = 1.5$. If the alternative is $H_a : \mu > 100$, the P-value is

$$1 - \Phi(1.5) = 0.0668$$

- If $\alpha = 0.05$, H_0 cannot be rejected.

- Now consider the case of $H_a : \mu \leq \mu_0$
- For simplicity, consider the case $\alpha = 0.05$. Then,

$$P - \text{ value } = P(Z \leq z | H_0) = \Phi(z)$$

  which is the area under Z-curve to the left of z.
- If e.g. $H_a : \mu < 100$, and $z = 22.75$, the P-value is $\Phi(22.75) = 0.0030$
- Thus, we reject H_0 at the level of 0.05.
- This test is called a *lower-tailed test*.

- Finally, let $H_a : \mu \neq \mu_0$ - a *two-tailed test*
- Let the null value be 100 and $x = 103$ that results in observed $z = 1.5$
- In practice, any $z \geq 1.5$ and $z \leq -1.5$ are more contradictory to H_0 than the observed z
- The P-value is

$$P(Z \geq 1.5 \text{ or } Z \leq -1.5) = 1 - \Phi(1.5) + \Phi(-1.5)$$
$$= 2[1 - \Phi(1.5)] = .1336$$

- In other words, P-value as a function of z is $2[1 - \Phi(|z|)]$.

Summary

- Let $H_0 : \mu = \mu_0$; define the test statistic $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$.
  1. $H_a : \mu > \mu_0$ has P-value $P(Z \geq z | H_0 \text{ is true })$ is called an *upper-tailed test*
  2. $H_a : \mu < \mu_0$ has the P-value $P(Z \leq z | H_0 \text{ is true })$ and is called an *lower-tailed test*
  3. $H_a : \mu \neq \mu_0$ has the P-value $P(Z \geq z \text{ or } Z \leq -z | H_0 \text{ is true })$ and is called a *two-tailed test*

Recommended Steps for Testing Hypotheses about a Parameter

1. Identify the parameter of interest and describe it in the context of the problem situation.
2. Determine the null value and state the null hypothesis.
3. State the alternative hypothesis.
4. Give the formula for the computed value of the test statistic.
5. Compute any necessary sample quantities, substitute into the formula for the test statistic value, compute that value
6. Determine the P-value and compare it to the level of significance level α.

- The formulation of hypotheses (steps 2 and 3) should be done before examining the data.

Example

- A manufacturer of sprinkler systems used for fire protection in office buildings claims that the true average system-activation temperature is 130. Can we believe his claim?
- Parameter of interest is $\mu =$ true average activation temperature.
- $H_0 : \mu = 130$; $H_a : \mu \neq 130$.
- Test statistic is

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{x} - 130}{1.5/\sqrt{n}}$$

- With $n = 9$ and $\bar{x} = 131.08$,

$$z = \frac{131.08 - 130}{1.5/\sqrt{9}} = 2.16$$

- The P-value is $2[1 - \Phi(2.16)] = 0.0308$ and so we fail to reject H_0 at significance level 0.01.

Type II Error

- As an example, consider an upper-tailed test where we reject H_0 if $P - value$ is $\leq \alpha$
- H_0 is not rejected when $\bar{x} < \mu_0 + z_\alpha \cdot \sigma/\sqrt{n}$
- For a particular $\mu^{'} > \mu$, the probability of Type II error is then

$$\beta(\mu^{'}) = P(\bar{X} < \mu_0 + z_\alpha \cdot \sigma/\sqrt{n} | \mu = \mu^{'})$$
$$= P\left(\frac{\bar{X} - \mu^{'}}{\sigma/\sqrt{n}} < z_\alpha + \frac{\mu_0 - \mu^{'}}{\sigma/\sqrt{n}} | \mu = \mu^{'}\right)$$
$$= \Phi\left(z_\alpha + \frac{\mu_0 - \mu^{'}}{\sigma/\sqrt{n}}\right)$$

- Similar derivations can help us to derive Type II error probabilities for a lower-tailed test and a two-tailed test. Results can be summarized as follows:
  1. $H_a : \mu > \mu_0$ has the probability of Type II Error $\Phi\left(z_\alpha + \frac{\mu_0 - \mu^{'}}{\sigma/\sqrt{n}}\right)$
  2. $H_a : \mu < \mu_0$ has the probability of Type II Error $1 - \Phi\left(-z_\alpha + \frac{\mu_0 - \mu^{'}}{\sigma/\sqrt{n}}\right)$
  3. $H_a : \mu \neq \mu_0$ has the probability of Type II Error $\Phi\left(z_{\alpha/2} + \frac{\mu_0 - \mu^{'}}{\sigma/\sqrt{n}}\right) - \Phi\left(-z_{\alpha/2} + \frac{\mu_0 - \mu^{'}}{\sigma/\sqrt{n}}\right)$

Sample Size Determination

- Sometimes, we want to bound the value of Type II error for a specific value μ'.
- Consider the same sprinkler example. Fix α and specify β for such an alternative value. For $\mu' = 132$ we may want to require $\beta(132) = 0.1$ in addition to $\alpha = .01$.
- The sample size required for that purpose is such that

$$\Phi\left(z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right) = \beta$$

- Solving for n, we obtain

$$n = \left[\frac{\sigma(z_\alpha + z_\beta)}{\mu_0 - \mu'} \right]^2$$

and the same answer is true for a lower-tailed test

- For a two-tailed test, it is only possible to give an approximate solution. It is

$$n \approx \left[\frac{\sigma(z_{\alpha/2} + z_\beta)}{\mu_0 - \mu'} \right]^2$$

Example

- Let μ denote the true average tread life of a certain type of tire
- Test $H_0 : \mu = 30,000$ vs. $H_a : \mu > 30,000$, with $n = 16$, normal population distribution with $\sigma = 1500$.
- A test with $\alpha = .01$ requires $z_\alpha = z_{.01} = 2.33$
- The probability of making a Type II error when the alternative $\mu = 31,000$ is

$$\beta(31,000) = \Phi\left(2.33 + \frac{31,000 - 30,000}{1500/\sqrt{16}}\right) = \Phi(-.34) = .3669$$

Example

- In addition to $\alpha = .01$ also require that $\beta(31,000) = .1$
- First, find $z_{.1} = 1.28$
- $$n = \left[\frac{1500(1.28 + 2.33)}{30,000 - 31,000}\right]^2 = 29.32$$
- In practice, take $n = 30$

Large Sample Tests

- When the sample size is large, the z tests described earlier are modified to yield valid test procedures without requiring either a normal population distribution or a known σ.
- Let us assume $n > 40$. Then, the test statistic

$$Z = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

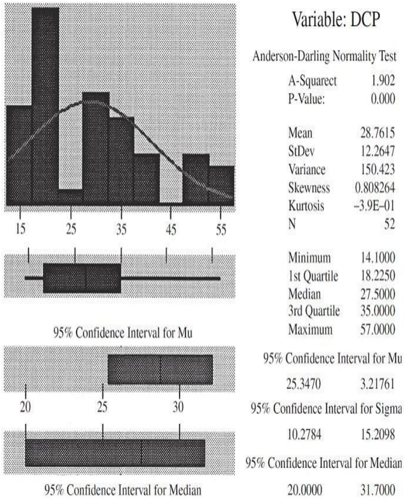
  is approximately standard normal
- The P-value will be computed exactly as before (but keep in mind that this will guarantee an approximate level of significance α only).

Example

- A dynamic cone penetrometer (DCP) is used for measuring material resistance to penetration (mm/blow) as a cone is driven into pavement or subgrade
- Suppose that for a particular application it is required that the true average DCP value for a certain type of pavement be less than 30.
- The pavement will not be used unless there is conclusive evidence that the specification has been met.

Example

Figure:



Example

- Note the lack of normality in the histogram!
- Hypotheses: $H_0 : \mu = 30$ vs. $H_a : \mu < 30$
- $z = \frac{\bar{x}-30}{s/\sqrt{n}}$
- $n = 52$, $\bar{x} = 28.76$, and $s = 12.2647$,

$$z = \frac{28.76 - 30}{12.2647/\sqrt{52}} = -.73$$

- The P-value is too large and H_0 cannot be rejected

A Normal Population Distribution

- When n is small, we can no longer invoke CLT as a justification for the large sample test
- Remember that for a normally distributed random sample $X_1, \ldots, X_n$, the statistic

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

  has a t distribution with $n - 1$ df
- Therefore, we have the test with $H_0 : \mu = \mu_0$ and a test statistic value $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$.

Summary of the Three Possible t-Tests

- If $H_a : \mu > \mu_0$, the P-value is the area under t_{n-1} curve to the right of the observed value t
- If $H_a : \mu < \mu_0$, the P-value is the area under t_{n-1} curve to the left of the observed value t
- If $H_a : \mu \neq \mu_0$, the P-value is twice the area under t_{n-1} curve to the right of $|t|$

Example

- The *Edison Electric Institute* publishes figures on the annual number of kilowatt hours expended by various home appliances.
- It is claimed that a vacuum cleaner expends an average of 46 kilowatt hours per year.
- Suppose a planned study includes a random sample of 12 homes and it indicates that VC's expend an average of 42 kilowatt hours per year with $s = 11.9$ kilowatt hours.
- Assuming the population normality, design a 0.05 level test to see whether VC's spend less than 46 kilowatt hours annually

- $H_0 : \mu = 46$ kilowatt hours and $H_a : \mu < 46$ kilowatt hours
- Assume $\alpha = 0.05$ and compute the observed value of the test statistic

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

- The value of the statistic is

$$t = \frac{42 - 46}{11.9/\sqrt{12}} = -1.16$$

- The P-value is the area under t_{11} curve to the left of -1.16
- Note that we fail to reject H_0

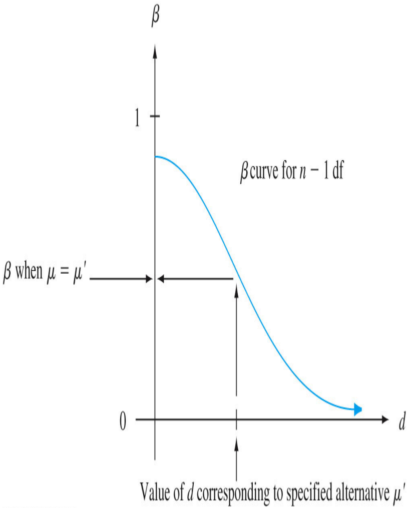
A β curve for the t-test

- It is much more difficult to compute the probability of the Type II Error in this case than in the normal case
- The reason is that it requires the knowledge of distribution of $T = \frac{\bar{X}-\mu_0}{S/\sqrt{n}}$ under the alternative H_a. To do it precisely, we must compute

$$\beta(\mu^{'}) = P(T < t_{\alpha,n-1} \text{ when } \mu = \mu^{'})$$

- There exist extensive tables of these probabilities for both one- and two-tailed tests.

Figure:


β
1
β curve for n − 1 df
β when μ = μ′
0
d
Value of d corresponding to specified alternative μ′

Calculating β

- First, we select $\mu^{'}$ and the estimated value for unknown σ. Then, we find an estimated value of $d = |\mu_0 - \mu^{'}|/\sigma$. Finally, the value of β is the height of the $n-1$ df curve above the value of d
- If $n-1$ is not the value for which the corresponding curve appears visual interpolation is necessary

- One can also calculate the sample size n needed to keep the Type II Error probability below β for specified α.
  1. First, we compute d
  2. Then, the point (d, β) is located on the relevant set of graphs
  3. The curve below and closest to the point gives $n-1$ and thus n
  4. Interpolation, of course, is often necessary

Large-Sample Tests

- Let p denote the proportion of individuals or objects in a population who possess a specified property; thus, each object either possesses a desired property (S) or it doesn't (F).
- Consider a simple random sample $X_1, \ldots, X_n$. If the sample size n is small relative to the population size, the number of successes in the sample X has an approximately binomial distribution. If n itself is also large, both X and the sample proportion $\hat{p} = X/n$ are approximately normally distributed
- Large-sample tests concerning p are a special case of the more general large-sample procedures for an arbitrary parameter θ. We considered such a large-sample test before for the mean μ of an arbitrary distribution.

- Some basic properties of $\hat{p}$ are;
  1. Estimator $\hat{p}$ is unbiased: $E\,\hat{p} = p$.
  2. Second, it is approximately normal and its standard deviation (SD) is $\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$
  3. Note that $\sigma_{\hat{p}}$ does not include any unknown parameters. This is not always the case. It is enough to remember the large-sample test of the mean where $\sigma_{\hat{\mu}} = \sigma_{\bar{X}} = \sigma^2/n$ which is in general unknown unless σ^2 is specified.

- Let us consider first an upper-tailed test. It means having a null hypothesis $H_0 : p = p_0$ vs. an alternative $H_a : p > p_0$.
- Under the null hypothesis, we have $E(\hat{p}) = p_0$ and $\sigma_{\hat{p}} = \sqrt{p_0(1 - p_0)/n}$; therefore, for large n the test statistic

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}$$

  has approximately standard normal distribution
- The P-value is clearly, then $1 - \Phi(z)$ where z is the observed value of Z

- The lower-tailed test has the P-value $\Phi(z)$
- The two-tailed test has the P-value $=2[1-\Phi(|z|)$
- These tests are applicable whenever the normal approximation of the binomial distribution is reasonable: $np_0 \geq 10$, $n(1-p_0) > 10$.

Example

- 1276 individuals in a sample of 4115 adults were found to be obese as per their BMI. A 1998 survey based on self-assessment revealed that 20% of the adult population considered themselves obese. Does the most recent data suggest that the true proportion of obese adults is more than 1.5 times the percentage from the self-assessment survey?
- The null hypothesis here is that $p = 0.3$ and the alternative is $p > .30$
- Note that the rule of thumb is satisfied: $np_0 = 4115(.3) > 10$ and $n(1 - p_0) = 4115(.7) > 10$
- Thus, we use the large-sample test with

$$z = (\hat{p} - .3)/\sqrt{(.3)(.7)/n}$$

- The sample proportion is $\hat{p} = 1276/4115 = .310$ and the corresponding value of z is 1.40
- Check that the P-value is less than 0.1

Type II Error and sample size determination

- Type II Error probability can be computed exactly as before. If H_0 is not true, the true proportion $p = p^{'} \neq p_0$. Under $H_a : p = p^{'}$ we have Z is still approximately normal; however,

$$E(Z) = \frac{p^{'} - p_0}{\sqrt{p_0(1 - p_0)/n}}$$

and

$$V(Z) = \frac{p^{'}(1 - p^{'})/n}{p_0(1 - p_0)/n}$$

- The formulas for the type II error are very similar to what we saw before for the mean test. We only give the upper-tailed test formula ($H_a : p > p_0$)

$$\beta(p^{'}) = \Phi\left[\frac{p_0 - p^{'} + z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p^{'}(1-p^{'})/n}}\right]$$

and the lower-tailed test formula ($H_a : p < p_0$)

$$\beta(p^{'}) = 1 - \Phi\left[\frac{p_0 - p^{'} - z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p^{'}(1-p^{'})/n}}\right]$$

- Sample size formulas can also be easily derived. In the two-tailed case, the formula is approximate as before

Example

- A package delivery service advertises that at least 90% of all packages brought to its office by 9 A.M. are delivered by noon the same day. How likely is it that a level 0.1 test based on $n = 225$ packages will detect a departure of 10% from this goal?
- The null hypothesis is $H_0 : p = 0.9$ vs. $H_a : p < 0.9$.
- For $p' = 0.8$, we have

$$\beta(.8) = 1 - \Phi\left(\frac{.9 - .8 - 2.33\sqrt{(.9)(.1)/225}}{\sqrt{(.8)(.2)/225}}\right) = 0.228$$

- Thus, the probability of type II error under the alternative $p' = 0.8$ is about 23%.

Small Sample tests

- These are test procedures for proportions when the sample size n is small. They are based directly on the binomial distribution rather than the normal approximation.
- Consider the alternative hypothesis $H_a : p > p_0$ and let X be yet again the number of successes in the sample size n.
- The P-value is

$$
\begin{aligned}
& P(X \geq x \text{ when } X \sim Bin(n; p_0)) \\
& = 1 - P(X \leq x - 1 \text{ when } X \sim Bin(n; p_0)) \\
& = 1 - B(x - 1; n, p_0)
\end{aligned}
$$

- It is usually not possible to get the test whose probability of Type I value is exactly α
- To compute the Type II error for an alternative $p^{'} > p_0$, we first note that $X \sim Bin(n, p^{'})$ if the alternative is true. Then,

$$\beta(p^{'}) = P(X < c_\alpha \text{ when } X \sim Bin(n, p^{'})) = B(c_\alpha - 1; n, p^{'})$$

  Note that this is a result of a straightforward binomial probability calculation

Example

- A builder claims that heat pumps are installed in 70% of all homes being constructed today in the city of Richmond, VA. Would you agree with this claim if a random survey of new homes in this city shows that 8 out of 15 had heat pumps installed? Use a 0.1 level of significance.
- $H_0 : p = 0.7$ vs. $H_a : p < 0.7$ with $\alpha = 0.10$
- The test statistic is $X \sim Bin(15, 0.7)$

- We have $x = 8$ and $np_0 = (15)(0.7) = 10.5$. Thus, the P-value is

$$P(X \leq x) = B(x; 15, 0.7)$$

which is equal 0.131 when $x = 8$.

Example

- A plastics manufacturer has developed a new type of plastic trash can and proposes to sell them with an unconditional 6-year warranty.
- To see whether this is economically feasible, 20 prototype cans are subjected to in accelerated life test to simulate 6 years of use.
- The proposed warranty will be modified only if the sample data strongly suggests that fewer than 90% of such cans would survive the 6-year period.

Example

- Let p denote the proportion of all cans that survive the accelerated test. The relevant hypotheses are $H_0 : p_0 = .9$ versus $H_a : p_0 < .9$.
- A decision will be based on the test statistic X, the number among the 20 that survive.
- Because of the inequality in H_a, any value smaller than the observed value x is more contradictory to H_0 than is x itself. Therefore,

$$P - \text{ value } = P(X \leq x|H_0) = B(x; 20, .9)$$

- Note: $B(15, 20, .9) = 0.043$ and $B(16; 20, .9) = .133$. Thus, 0.043 is the closest to 0.05 achievable significance level

Statistical vs. Practical Significance

- Consider testing $H_0 : \mu = 100$ vs. $H_a : \mu > 100$ where $\sigma = 10$
- Suppose the true value is $\mu^{'} = 101$; if the sample size is large, $\bar{x}$ will be close to $\mu^{'}$ most of the time and H_0 will be rejected with high probability...But what if it is not significant practically?!
- Be careful in interpreting evidence when the sample size is large, since any small departure from H_0 will almost surely be detected by a test (*statistical significance*), yet such a departure may have little *practical significance*.

Statistical vs. Practical Significance

n	P-value when $\bar{x} = 101$	$\beta(101)$ for level .01 test
25	.3085	.9664
100	.1587	.9082
400	.0228	.6293
900	.0013	.2514
1600	.0000335	.0475