Lecture 15

Part 3 Estimation and Hypothesis Test

Continue our talk on hypothesis test

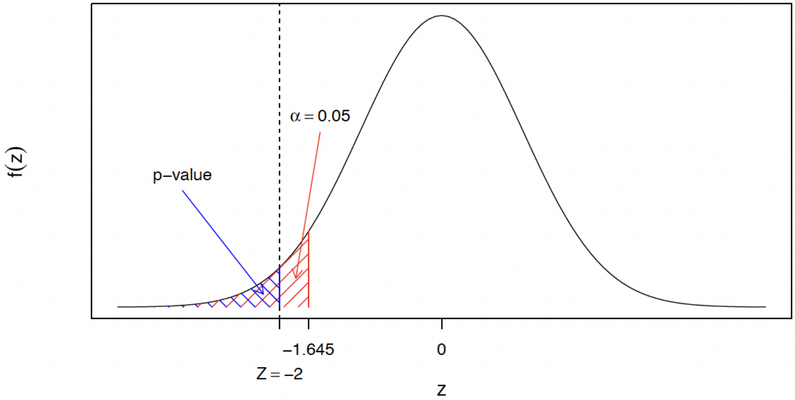
Importance of p-values

- How can we use p-values to conduct a hypothesis test?
- A very small p-value means that the observed test statistic was very extreme.
- So very small p-values should result in the rejection of H_0.

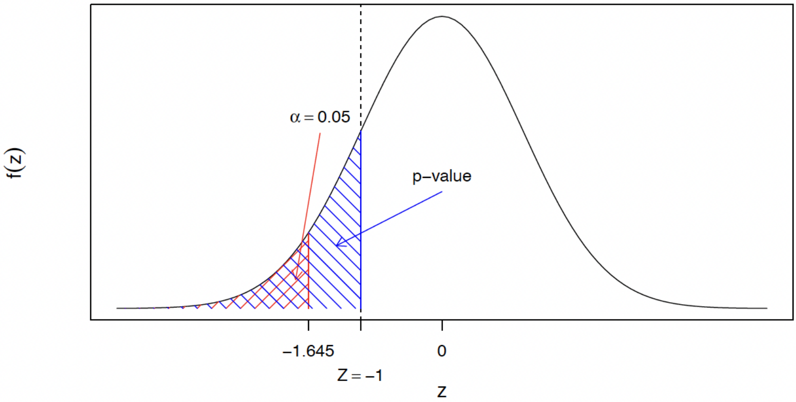
Importance of p-values

- But how small is small?
- We can perform the hypothesis test by comparing the p-value to the significance level α:
  - If the p-value is less than α, we reject H_0.
  - If the p-value is greater than α, we fail to reject H_0.
- Why?
- Suppose we test $H_0 : \mu = \mu_0$ vs $H_1 : \mu < \mu_0$ at $\alpha = 0.05$.

Importance of p-values



Importance of p-values



Importance of p-values

- Remember that rejection regions were calculated by finding the critical values that cut off $100\alpha\%$ in the extreme tail (one-tailed) or tails (two-tailed).
- If the p-value is less than α, then the test statistic must lie within the rejection region, which means we reject H_0.
- If the p-value is greater than α, then the test statistic must not lie within the rejection region, which means we fail to reject H_0.

Another Bowling Example

- Your lecturer claims to have a bowling average of 150 or higher.
- You play 10 games with him, and he scores an average of 140.
- Suppose that the standard deviation for bowling scores is 15.
- Given a 5% significance level, do you reject your lecturer's claim?

Hypotheses and Test Statistic

1. State the hypotheses:

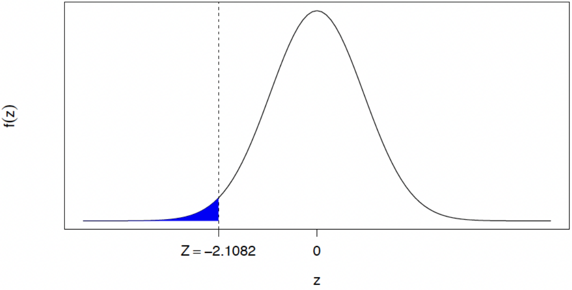
$$H_0 : \mu = 150$$
$$H_1 : \mu < 150$$

2. Standardize $\bar{X}$ to get the test statistic:

$$Z = \frac{140 - 150}{\frac{15}{\sqrt{10}}} = -2.1082$$

Decision Rule and Conclusion

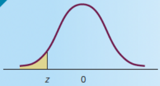
1. The p-value is: $P(Z < -2.1082) = 0.0174$.



2. Since p-value $< \alpha$, i.e., $0.0174 < 0.05$, we reject H_0 and conclude my bowling average is less than 150.

Decision Rule and Conclusion

TABLE 3 Cumulative Standardized Normal Probabilities



$P(-\infty < Z < z)$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559

Using p-values

- The conclusions we draw from using the rejection region method and the p-value method are always *exactly* the same.
- When calculating p-values for two-tailed tests, remember that "more extreme" includes the situations when the test statistic is too small *or* too big. We care about *both* tails.
- Using p-values is a more elegant and flexible way of performing hypothesis tests than using rejection regions.

Confidence Intervals

- Confidence intervals for μ and standardized test statistics used in hypothesis tests for μ are both derived from the sampling distribution of $\bar{X}$.
- Let's take a brief moment to compare a $100(1-\alpha)\%$ confidence interval and a two-tailed hypothesis test at a significance level of α.

Hypothesis Tests and Confidence Intervals

- Confidence interval is given by:

$$\left(\bar{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

- We reject a two-tailed hypothesis test at a significance level of α if:

$$\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} < -z_{\frac{\alpha}{2}} \quad \text{or} \quad \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > z_{\frac{\alpha}{2}}$$

Hypothesis Tests and Confidence Intervals

- Starting with the lower rejection region and rearranging, we get:

$$\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} < -z_{\frac{\alpha}{2}}$$

$$\Rightarrow \quad \bar{X} - \mu_0 < -z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow \quad \mu_0 > \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

- We can perform a similar rearrangement for the upper rejection region.

Hypothesis Tests and Confidence Intervals

- Inspecting these carefully, it becomes clear that constructing a confidence interval is equivalent to performing a *two-tailed* hypothesis test.
- If the value of μ_0 in the null hypothesis falls *outside* the calculated confidence interval, we *reject* the null hypothesis.
- If it falls *inside* the calculated confidence interval, we *do not reject* the null hypothesis.

Type I and Type II Errors

- Type I error occurs when we reject the null hypothesis when it is true.
  - $P(\text{Type I error}) = \alpha$, the significance level of the test.
- Type II error occurs when we fail to reject the null hypothesis when it is not true.
  - $P(\text{Type II error}) = \beta$.
  - The **power** of a test is defined to be $1 - \beta$.

Probability of a Type II Error

- Calculating $P(\text{Type II error})$ involves finding the probability of not rejecting H_0 (i.e., the test statistic not falling in the rejection region) when H_1 is true (i.e., under the "alternative distribution").
- To do this we need to know:
  - The rejection region.
  - The true value of the parameter under H_1.

Probability of a Type II Error

- $P(\text{Type II error})$ can generally be calculated using the following steps:

1. Based on α and the hypotheses, determine the rejection region in terms of the standardized test statistic.
2. Works backwards to determine the rejection region in terms of the *unstandardized* test statistic.
3. Calculate the probability of not rejecting H_0, under H_1, by re-standardizing using the true value of the parameter.

Example

- You want to test the claim that the average house price in your suburb is not below \$400,000. The average of the most recent five sales is equal to \$420,000. The standard deviation of house prices is \$20,000.
- Given that the true average house price is in fact \$390,000, calculate the probability of a type II error and the power of the test, at a 5 percent significance level.

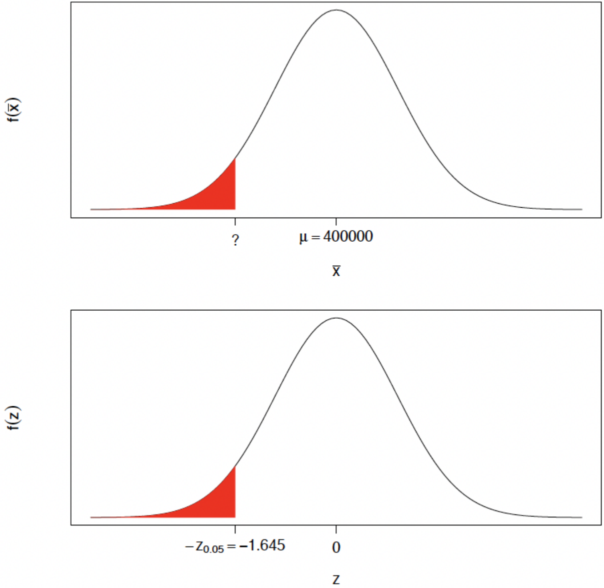
Hypotheses

- Note that the claim given in the question is $\mu \geq 400\,000$. So the hypotheses are:

$$H_0 : \mu = 400\,000$$
$$H_1 : \mu < 400\,000$$

- This is a one-tailed (lower-tailed) test.

Null Distribution and Rejection Region

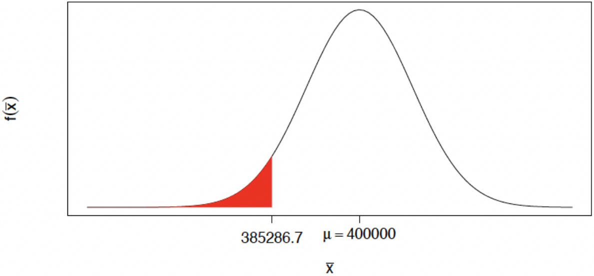


Null Distribution and Rejection Region

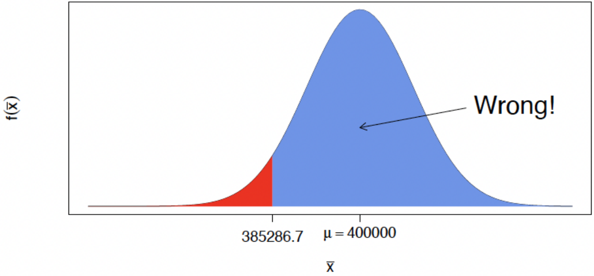
- We need to convert the rejection region in terms of the standardized Z-statistic, i.e., $Z < -1.645$, into a rejection region in terms of the unstandardized $\bar{X}$:

$$
\begin{aligned}
Z &< -1.645 \\
\Rightarrow \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} &< -1.645 \\
\Rightarrow \frac{\bar{X} - 400\,000}{\frac{20\,000}{\sqrt{5}}} &< -1.645 \\
\Rightarrow \bar{X} &< 385\,286.7
\end{aligned}
$$

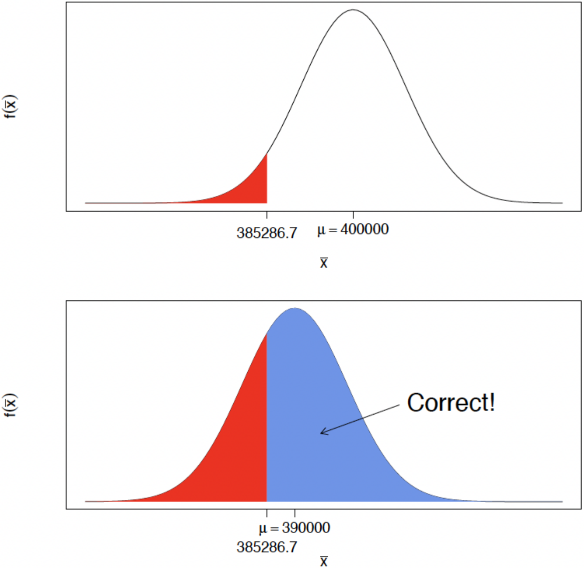
Null Distribution and Rejection Region



Probability of a Type II Error



Probability of a Type II Error



Probability of a Type II Error

- To calculate $P(\text{Type II error})$ we need to standardize with respect to the true value of μ under H_1, i.e., $\mu = 390\,000$:

$$
\begin{aligned}
P(\text{Type II error}) &= P(\bar{X} > 385\,286.7 \mid H_1 \text{ is true}) \\
&= P\left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} > \frac{385\,286.7 - 390\,000}{\frac{20\,000}{\sqrt{5}}}\right) \\
&= P(Z > -0.5270) \\
&= 1 - P(Z < -0.5270) \\
&= 0.7019
\end{aligned}
$$

Probability of a Type II Error

- So if the true average house price is actually \$390 000, then the probability of making a type II error is $\beta = 0.7019$, which is relatively high.
- This means the power of the test is equal to $1 - \beta = 0.2981$.
- Note that the $\beta = P(\text{Type II error})$ depends on both α and the true value of μ under H_1.

Hypothesis Test for μ when σ^2 is Unknown

Hypothesis Test for μ when σ^2 is Unknown

- When we know σ^2, hypothesis tests for μ use the standardized Z-statistic as the test statistic:

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

- The null distribution of Z was $N(0,1)$ (for large n), so we can determine rejection regions and calculate p-values by looking up the z-tables.
- What happens if we don't know σ^2?

We Often Don't Know the Population Variance σ^2

- In many real-world scenarios, the population variance is unknown
- Examples:
  - Variance of height of all the OU students.
  - Medical research: Testing a new drug's effectiveness
  - Quality control: Assessing product consistency in manufacturing
  - Social sciences: Studying income distribution in a large population

Hypothesis Test for μ when σ^2 is Unknown

- We can estimate σ^2 using the sample variance s^2.
- So let's replace σ in the Z-statistic with s:

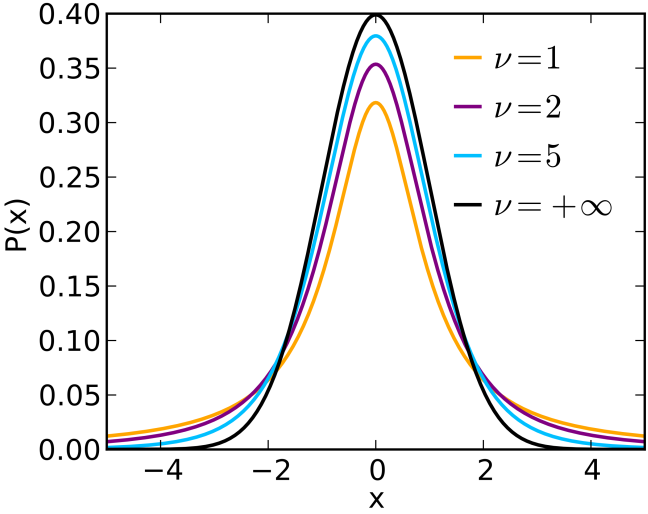
$$T = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}}$$

- Under H_0, this "T-statistic" no longer has a $N(0,1)$ distribution, so we can't perform the test by looking up the z-tables.
- But if we knew the null distribution of T, we could still perform the test...

t-distribution

- The null distribution of this T-statistic is the t-distribution with $n-1$ degrees of freedom.
- The t-distribution is a special continuous distribution:
  - It's symmetric about 0 and bell shaped, just like the standard normal distribution.
  - It has one parameter called the degrees of freedom, and as this increases, the t-distribution approaches the standard normal distribution.
  - It has a greater variance than the standard normal distribution.

t-distribution



Let's use the Bowling Example

- Your lecturer claims to have a bowling average of 150 or higher.
- You play 10 games with him, and he scores an average of 140.
- The population standard deviation of bowling scores is unknown. However, for these 10 games, you calculate the sample standard deviation to be 13.7.
- Given a 5% significance level, do you reject your lecturer's claim?

Hypotheses and Test Statistic

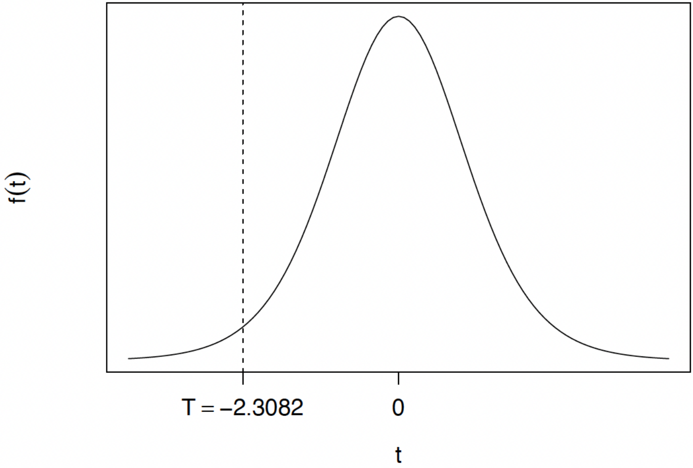
- Hypotheses:

$$H_0 : \mu = 150$$
$$H_1 : \mu < 150$$

- Test statistic:

$$T = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{140 - 150}{\frac{13.7}{\sqrt{10}}} = -2.3082$$

Null Distribution

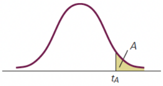


Decision Rule

- We need to determine the rejection region.
- That is, we must find the critical value that cuts off 5% in the lower tail of a t-distribution with $n - 1 = 9$ degrees of freedom.
- From the symmetry of the t-distribution about 0, the rejection region is given by $T < -1.833$.

Decision Rule

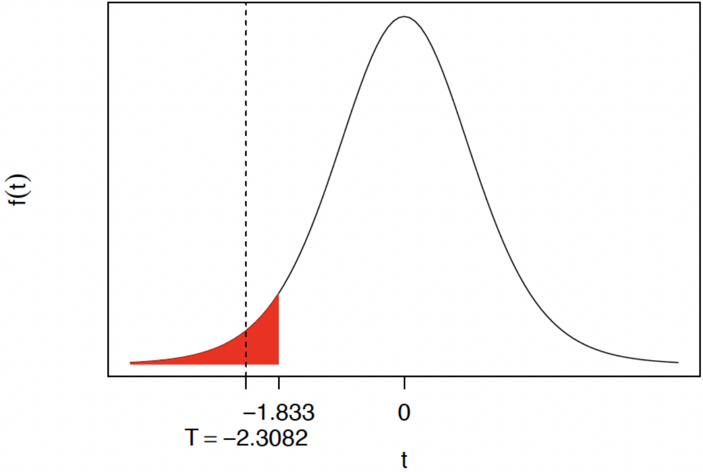
TABLE **4**
Critical Values of the Student *t* Distribution



Degrees of Freedom	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787

Conclusion

- Since $-2.3082 < -1.833$, we reject H_0 and conclude that my bowling average is less than 150.



Confidence Interval

- We know that calculating a confidence interval is the same as performing a two-tailed hypothesis test.
- A $100(1-\alpha)\%$ confidence interval for μ when σ^2 is unknown is given by:

$$\left(\bar{X} - t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}, \bar{X} + t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}\right)$$

- NB: $t_{\frac{\alpha}{2},n-1}$ is the value that cuts off an area of $\frac{\alpha}{2}$ in the upper tail of the t-distribution with $n-1$ degrees of freedom.

Hypothesis Test for p

- Suppose we are now interested in making inferences about a population proportion p.
  - For example, the population proportion of people who prefer Coke over Pepsi.
- How would we test the following hypotheses?

$$H_0 : p = p_0$$
$$H_1 : p(\neq, <, >)p_0$$

Hypothesis Test for p

- We can estimate a population proportion by the sample proportion:

$$\hat{p} = \frac{X}{n}$$

  where X is the number of successes in the sample and n is the sample size.

- From previous topics we know that, for large n, the sampling distribution of $\hat{p}$ is:

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

Hypothesis Test for p

- To perform hypothesis tests for p, we can then use the Z-statistic obtained by standardizing $\hat{p}$:

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

- Since $Z \sim N(0,1)$ under H_0, we can then proceed as we normally do using the z-tables.

Confidence Interval

- A $100(1-\alpha)\%$ confidence interval for a population proportion p is given by:

$$\left(\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

- Note that inside the square root we replaced p_0 with $\hat{p}$, since in an estimation setting we don't know the true value of p and there is no hypothesized value p_0.

Example

- The number of operators at a call center depends on the level of use.
- At random times over a brief interval, the call center is checked to see whether all operators are busy so as to assess the proportion of time all operators are busy.
- An additional operator will be employed if there is sufficient evidence that all operators are busy more than 80% of the time.

Example

- Twenty-five random times are chosen for checking and it was found that all operators were busy 96% of the time.
- Would you recommend employing an additional operator in the call center?
- Population parameter of interest is p, the proportion of time that all operators are busy.
- We know that $n = 25$ and $\hat{p} = 0.96$.

Hypotheses and Test Statistic

- Hypotheses:

$$H_0 : p = 0.8$$
$$H_1 : p > 0.8$$

- Test statistic (assume H_0 is true):

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.96 - 0.8}{\sqrt{\frac{0.8(1-0.8)}{25}}} = 2$$

Decision Rule and Conclusion

- Decision rule:
  - Calculate the p-value from the z-tables:

$$p\text{-value} = P(Z > 2) = 0.0228$$

- Conclusion:
  - Since $0.0228 < 0.05$, we reject H_0 at the 5% significance level.
  - We would conclude there is strong evidence that the proportion of time all operators are busy is greater than 80% and recommend that a new operator be employed.