9 Hypothesis Tests

Inference requires a prior "tentative assumption".
We begin by making a tentative assumption about a population parameter θ, called the Null Hypothesis, denoted by H_0.

- ▷ How much is the average rate of spread of Corona virus across countries?
- ▷ Lets say the average rate of spread is 22 days i.e. $\mu = 22$

We also think about the alternative hypothesis: is the opposite of the null, denoted by H_a

- ▷ Opposite 1: $\mu < 22$ or $\mu > 22$
- ▷ Opposite 2: $\mu \neq 22$

Hypothesis testing: tells you statistically which statement is true and which is false.

Consider an example based on this simple observation – I crossed the road after a black cat, had a bad day

- ▷ Hypothesis - cross the road after a black cat, you will have a bad day
- ▷ Test Hypothesis - collect more information about *that* day of other people who crossed the *same* road after the black cat
- ▷ Result - Reject or accept the hypothesis

Problem: It is not always obvious how the null and alternative hypotheses should be formulated.

- ▷ Collect a sample → use your knowledge of point and interval estimates

9.1 Developing Null and Alternative Hypotheses

Sometimes it is easier to think about an alternative hypothesis

- In such cases you want to reject the null such as: crossing the road after a black cat causes a bad day.

Example: Honda Jazz

Consider the case of Honda Jazz. They hire Yu to test whether their new i-VTEC engine has a better mileage than other cars in this segment. Yu collects a sample of different Honda Jazz cars, which are driven by automated robots.

What is your hypothesis to test the mileage of Jazz?

- Do you think the mileage would be different across these robots?
- If they drive on the exact same road, then the mileage should be identical.
- What is the point of testing?

Yu test to determined if it can be concluded that the new system provides more than 24 miles per gallon. Null hypothesis for Jazz is

- population mean miles per gallon $H_0 : \mu \leq 24$
- Alternative: $H_a : \mu > 24$

Conclusion: if the sample results lead to the conclusion to reject H_0, the inference can be made that $H_a : \mu > 24$ is true.

Another Example:
Let's say you work for the Republican party

- How many would work for them? and Why?

The party hires you to find out if they will win the election

- One reasonable way to put this question: will 51% of the electorate will vote for them

Formulate a hypothesis that puts this question in statistical terms!

- $H_0 : p \geq 0.51$ (what's p?)
- $H_a : p < 0.51$

Formulation: divides the problem into two: one with the possibility of winning, another with possibility of *not winning* (need not be losing)

Alternative way (or need) of formulating a null hypothesis:

- Belief or (an assumption) that a statement about the value of a population parameter is true
- Testing: to challenge the assumption or to determine whether the belief is correct or not.

Another Example:
Claim: a bottle of Coke has less than 0.0005 mg/l of pesticide as per the Food Safety and Standards Authority of India (FSSAI)

- Whats the first step? Collect a sample..
- $H_0 : \mu \leq 0.0005 mg/l$ and the alternative is $H_a : \mu > 0.0005 mg/l$
- Collect a sample to find that $0.0180 mg/litre$ is the actual and hence reject the null.

(a) In this chapter we learn about hypothesis on population mean and population proportions

(b) We distinguish between the types of Hypothesis

$$
\begin{array}{lll}
H_0 : \mu \geq \mu_0 & H_0 : \mu \leq \mu_0 & H_0 : \mu = \mu_0 \\
H_a : \mu < \mu_0 & H_a : \mu > \mu_0 & H_a : \mu \neq \mu_0
\end{array}
$$

First two are called one tailed tests, Last is called the two tailed test.

9.2 Type I and Type II Errors

We need to test the hypothesis, once you have formulated it. Unfortunately, the correct conclusions are not always possible.

Type II error: is when we accept H_0 an incorrect conclusion, when H_a was actually true

- He is a poltergeist since he could fly and wore no shoes, but it actually turns out that he is superman!

Type I error: is we reject H_0 when its actually true

- Think of the Murder on the Orient Express (1974): A Type I error would be when Hercule Poirot (the detective from famous Agatha Christie novels) would absolve the actual killer of crime.

TABLE 9.1 Errors and Correct Conclusions in Hypothesis Testing

		Population Condition	
		H_0 True	H_a True
Conclusion	Accept H_0	Correct Conclusion	Type II Error
	Reject H_0	Type I Error	Correct Conclusion

Vatsalya's Example

1. Ask a question with at-least two possible answers
2. Design a test out of the sample you collected
3. Base your decision on the test
    - ▷ There is a possibility that you commit either Type 1 or 2 error and you would base your decisions such that you minimize the possibility

4. Consider the Neanderthals and their dilemma

   - ▷ Not sure if there is a Lion outside the den (cave men!)
   - ▷ H_0: There is no lion outside
     H_a: There is lion outside.
   - ▷ Commit Type 1 error: you are scared and starving thinking there is lion i.e. accept H_a, but actually there is no lion H_0 is true
   - ▷ Commit Type 2 error: you are brave and reckless assuming there is no lion i.e. accept H_0; but actually H_a is true

Example: Honda Jazz and Testing: Their hypothesis on the average miles per gallon implied them

$$H_0 : \mu \leq 24 \qquad H_a : \mu > 24$$

Type 1 error: Reject H_0 when it is true, implies that the researchers claim the new system improves the miles per gallon falsely

Type 2 error: Accept H_0 when it is false, i.e. they have actually built a better car, but they aren't marketing it.

The probability of rejecting $H_0 : \mu \leq 24$, when it is true (as an equality i.e. $\mu = 24$) is the "level of significance" in our Jazz example.

Definition 9.1. Level of Significance α is the probability of making a Type 1 error when the null hypothesis is true as an equality

Common values are 0.05 and 0.01, hypothesis tests which control only Type 1 error called 'significance tests'

Definition 9.2. Inference Limits with Significance Tests: If the sample data are consistent with the null hypothesis H_0, we will follow the practice of concluding "do not reject H_0" this conclusion is preferred over "accept H_0" because the conclusion to accept H_0 puts us at risk of making a Type II error.

9.3 Population Mean with Known σ

Known population standard deviation, σ: if information from previous surveys is made available

- Able to make accurate inference if population this population is distributed normally, since we know the area under a standard normal curve.
- Not Normal: then need large sample to improve accuracy.

One-tailed tests are when we use an imprecise null i.e. H_0 gives us a range

Lower Tail Test	Upper Tail Test
$H_0 : \mu \geq \mu_0$	$H_0 : \mu \leq \mu_0$
$H_a : \mu < \mu_0$	$H_a : \mu > \mu_0$

Example: *Hilltop coffee and FTC*

The Federal Trade Commission (FTC) periodically conducts statistical studies designed to test the claims that manufacturers make about their products.

Hilltop Coffee: a coffee container labeled $3\,lb$, might have more or less - packaging errors bound to exist

FTC: as long as population average weight (population here is all the coffee packets produced in a year) is 3lb, consumers are paying fair price. Note that the FTC will not check each and every coffee container, but only a select random sample.

FTC can set up a significance test to check these claims - a lower tail test in particular because it does not bother if Hilltop coffee fills more than 3lbs in its containers.

The FTC will undertake the following steps. We will think through the problem as if we are officers from FTC:

Step 1: Developing a null hypothesis

If the population average weight for their coffee cans is at least 3 lb, then their package claims are correct i.e Null H_0

Hypothesized value for the population average is: $\mu_0 = 3$

$$H_0 : \mu \geq 3 \qquad H_a : \mu < 3$$

Sample data indicates that H_0 can be rejected then H_a could be true in significant cases, $\therefore$ violation has occurred.

How much lower than 3 can $\bar{x}$ be, before we declare that there is label violation? Risk of wrongly accusing Hilltop coffee, i.e. making Type I error of rejecting null when its true

Step 2: Decide the level of significance, α

α is the probability of making a Type-I error by rejecting H_0 when the null hypothesis is true as an equality when $\mu_0 = 3$

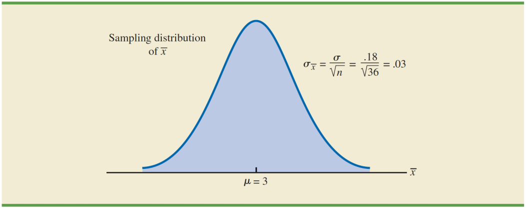
The commission says a maximum of 1% chance rejecting their claims when it is true is acceptable.

Step 3: Constructing a Test Statistic

We obtain a sample of 36 containers, and from previous such tests is 2018 and 2019, we know $\sigma = 0.18$
Assume population distribution is normal and standard error is given by

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.18}{\sqrt{36}} = 0.03$$

FIGURE 9.1 SAMPLING DISTRIBUTION OF $\bar{x}$ FOR THE HILLTOP COFFEE STUDY WHEN THE NULL HYPOTHESIS IS TRUE AS AN EQUALITY ($\mu = 3$)



Transform sample point estimator to a test statistic by demeaning (subtracting by μ_0) and standardizing (dividing by $\sigma_{\bar{x}}$)

Sampling distribution of the test statistic is given by

$$z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}$$

Use knowledge about standard normal distribution and talk about the corresponding value of $\bar{x}$

- If $z = -3$, we know its 3 standard errors below mean (z's mean is 0), so its probability is 0.00135.
- What is true for $z = -3$ is also true for $\bar{x}$ hence obtained from the sample given hypothesized value and standard error.

Aside: Z-value

Implies the value of sample mean (demeaned with hypothesized value μ_0) and standardized with known population error (now the sd of z is 1)

- Remember that $z \sim N(0, 1)$
- Then, a value of $z = -1$ implies that $\bar{x}$ is one standard deviation below the hypothesized value of the mean, whatever that may be.
- We look up the standard normal probability table to find a lower tail probability when $z = -3.0$
  - This implies a probability of 0.00135
- Inference: the probability of obtaining a value of z that is three or more standard errors below the mean is 0.00135
  - Such a result is unlikely if the null hypothesis is true

Definition 9.3. Test Statistic: when population σ is known and the population distribution of the weights of coffee container is normal then

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

Step 4: Calculate the p-value
How small should the z-value be before we reject the null hypothesis?

Definition 9.4. p-value: is a probability that provides a measure of the evidence against the null hypothesis provided by the sample, when the null is true. [Check this video]

Lower tail test: is a p-value such that the probability of obtaining a value for the test statistic smaller than the its sample counterpart.

We compute the test statistic, z and find the corresponding p-value looking up the table. Then we have to decide whether this is small enough based on α

9.3.1 p-value Method

Step 4: Calculate the p-Value

Hilltop Coffee sample collected by the FTC: $\bar{x} = 2.92$, $\sigma = 0.18$ and $n = 36$

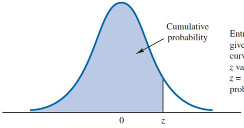
- Is $\bar{x}$ small enough to conclude that Hill Top Coffee has been violating the packaging requirements?
- Challenge is to claim about population from this particular sample.

Compute the test statistic:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = -2.67$$

p-value is the area in z-table to the left of $z = -2.67$. Written as: $p(z \leq -2.67) =?$ (look up table)

TABLE 1 CUMULATIVE PROBABILITIES FOR THE STANDARD NORMAL DISTRIBUTION (*Continued*)



Entries in the table give the area under the curve to the left of the z value. For example, for $z = 1.25$, the cumulative probability is .8944.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.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	.9279	.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

Obtained $p(z \leq 2.67) = 0.9962$. How do I calculate $p(z \leq -2.67)$? Since Standard normal is symmetric: $p(z \geq 2.67) = p(z \leq -2.67)$

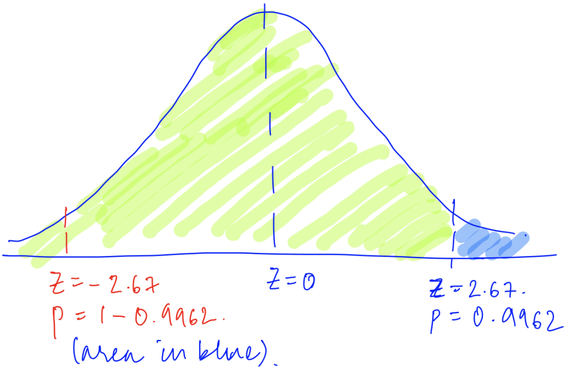
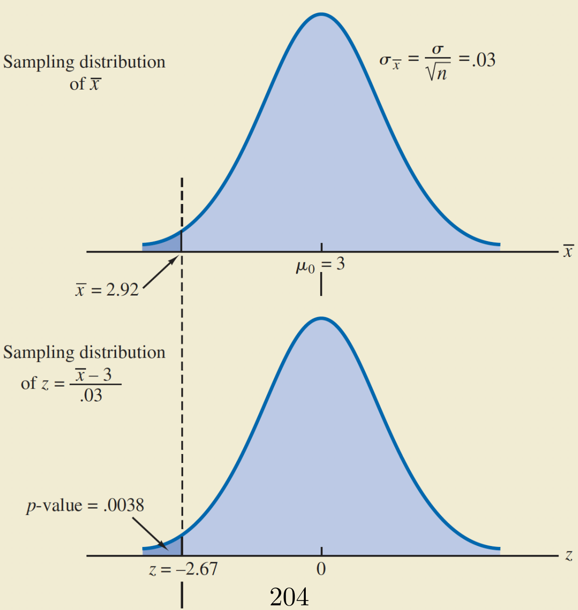


FIGURE 9.2 p-VALUE FOR THE HILLTOP COFFEE STUDY WHEN $\bar{x} = 2.92$ AND $z = -2.67$



Step 4: Calculate the p-Value when $\alpha = 0.01$ was decided by the FTC director.

Result: The sample of 36 coffee cans in the Hilltop coffee study resulted in a p-value $= 0.0038$, which means that the probability of obtaining a value of $\bar{x} = 2.92$ or less when the null hypothesis is true as an equality is 0.0038.

Definition 9.5. Rejection Rule using P-Value We reject H_0 iff p-value $\leq \alpha$

Inference: Given the population is normal and that $\sigma = 0.18$, there is a 0.3% chance we commit a Type-I error i.e. reject the null when its true

Hypothesis Result: FTC finds sufficient statistical evidence to reject the null hypothesis at *even* 1%, when it is true.

- It rejects the null at 0.3%, and hence does it at 0.4%, 0.5% and so on..
- Since p = 0.00397, we say that this is the *observed level of significance*.

9.3.2 Critical value approach

Step 4A: Ascertain the Critical Value: approach requires that we first determine a value for the test statistic deemed the "critical value".

Instead of focusing on the probability of committing Type-1 error, we focus on how much should our test statistic be...

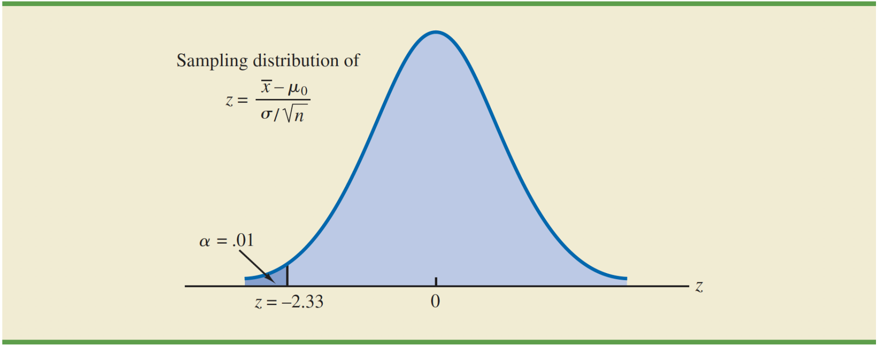
- ▷ The critical value is the largest value of the test statistic that will result in the rejection of the null hypothesis.
- ▷ instead of relying on a p, ask what is the lowest z value that we can tolerate corresponding to an α

Hilltop Coffee example: For a $\alpha = 1\%$ rejection of labels what is the value of the z-statistic?

- What is the value of z which corresponds to a lower tail probability of 1% ?
  - z = -2.33 provides a probability of $0.0099 \approxeq 0.01$
- If test statistic is less than of equal to -2.33, the corresponding p-value will be less than or equal to 0.01.
  - We then reject the null hypothesis.
  - Our z = -2.67 is less than -2.33 and hence we can say that CCD is underfilling cans and supplying consistently less than their claimed 3 lb.

Definition 9.6. Rejection Rule: Critical Value Approach Reject H_0 iff $z \leq -z_\alpha$ where $-z_\alpha$ is the critical value; that is, the z value that provides an area of a in the lower tail of the standard normal distribution.

FIGURE 9.3 CRITICAL VALUE = −2.33 FOR THE HILLTOP COFFEE HYPOTHESIS TEST



P-value is preferred: the advantage of the p-value approach is that the p-value tells us how significant the results are (the observed level of significance)

Definition 9.7. Summarizing the One Tailed Test

- Choose the critical value
- Compute the value of the test statistic when σ is known and population is assumed to be normally distributed.
- Lower tail test: using the standard normal distribution, compute the probability that z is less than or equal to the value of the test statistic (area in the lower tail).
- Upper tailed: is nothing but 1 - lower tail probability, so all problems can be reduced to a lower tailed problem.

9.3.3 Two Tailed Tests

Kookaboora's Cricket Balls Example: need to have the average weight of 163 gms for its white balls

Why is the weight of the ball so important? Heavier implies more T20 6's and lighter implies too much swing and hence less 6's.

Need exactly $\mu = 163$gms, hence they plan to test their production process. The quality control team would draw a sample of

- n = 50 balls
- Alternative hypothesis is $\mu \neq 163$

Choose a level of significance of 0.05 i.e. 5% chance of committing Type 1 error, with $\sigma = 12$ from tests last year

Standard error of sample average is

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{50}} = 1.7$$

Hypothesized population mean i.e null is $\mu_0 = 163$

- Sample average weight is $\bar{x} = 166$. Is this sample average different from the hypothesized value so that we reject the null at 5%?

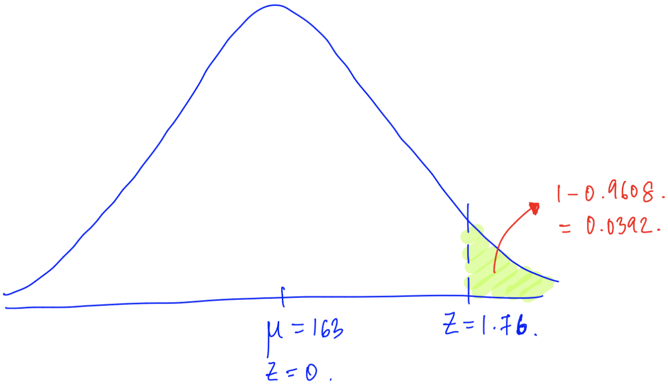
P-Value Approach

- Two tails: values of the test statistic in either tail provide evidence against the null hypothesis.

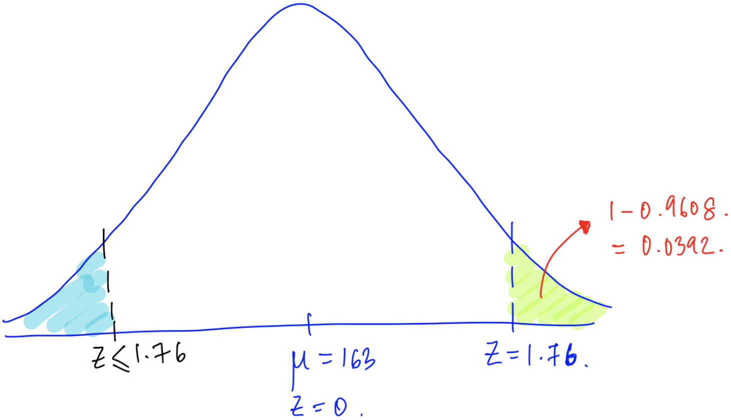
- Test statistic: therefore compute the test statistic for one tail and double the probability to include both tails.

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{166 - 163}{1.7} = 1.764$$

Probability of obtaining the test statistic at least as *unlikely* as $z = 1.764$



$P(Z \geq 1.76) = (0.0392).$

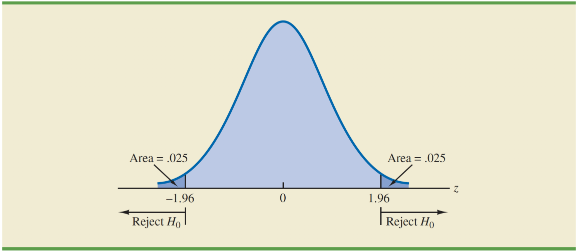


$P(Z \geq 1.76) = (0.0392)$.

$P(Z \leq -1.76) = 0.0392$ using $Z \sim N(0,1)$.

For two tails: we add the probability twice since the unlikely z-score occurs on both sides of the distribution: i.e. p-value is 0.0784

- Significance level: of 5% with the $p - \text{value} > \alpha$
- Cannot reject the null hypothesis at 5% level of significance.
    - We are probable to reject the null 7% of the times when it is true, i.e. the balls were the correct weight but we rejected it based on our sample



Critical-Value Approach: With $\alpha = 0.5$, we attribute the probability both sides equally

- and find the value of z in the table which corresponds to 1- 0.025 = 0.975
- Z = 1.96
- Two tailed rejection rule using the value of the test statistic is now iff $z \leq -1.96$ or $z \geq 1.96$
- Kookaboora eg: $z = 1.76 < 1.96$, cannot reject the null

Summarizing Hypothesis Tests with Known σ

TABLE 9.2 SUMMARY OF HYPOTHESIS TESTS ABOUT A POPULATION MEAN: σ KNOWN CASE

	Lower Tail Test	**Upper Tail Test**	**Two-Tailed Test**
Hypotheses	H_0: $\mu \geq \mu_0$ H_a: $\mu < \mu_0$	H_0: $\mu \leq \mu_0$ H_a: $\mu > \mu_0$	H_0: $\mu = \mu_0$ H_a: $\mu \neq \mu_0$
Test Statistic	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$
Rejection Rule: *p*-Value Approach	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$
Rejection Rule: Critical Value Approach	Reject H_0 if $z \leq -z_\alpha$	Reject H_0 if $z \geq z_\alpha$	Reject H_0 if $z \leq -z_{\alpha/2}$ or if $z \geq z_{\alpha/2}$

9.3.4 Interval Estimation and Hypothesis Testing

$(1-\alpha)\%$ Confidence interval: based on normality is given by

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

- Inference: we know that $100(1-\alpha)\%$ of the confidence intervals generated will contain the population mean and $100\alpha\%$ of the confidence intervals generated will not contain the population mean.

- Reject null hypothesis when its true with a probability $\alpha \simeq$ to conducting a two tailed test with α as level of significance...

Kookaburra's Testing Procedures

- Kookaburra's weight requirements imply the following null and alternative:

$$H_0 : \mu = 163 \qquad H_a : \mu \neq 163$$

- With a 50 ball sample, we obtain a sample average of $\bar{x} = 166$, with known population standard deviation as $\sigma = 12$
  - For a 5% confidence interval we obtain z = 1.96 as before
  - Interval implies:

$$166 \pm 1.96 \frac{12}{\sqrt{50}} \qquad 166 \pm 3.332$$

    or an interval from 162.668 to 169.332
  - Because the null lies between this range, we cannot reject the null through hypothesis testing

9.4 Population Mean with Unknown σ

- Since we do not have any previous studies to fall back on and obtain population σ, we rely on sample mean and standard deviation s, to infer characteristics of population and conduct hypothesis tests

- Instead of a standard normal distribution

  - We use a t-distribution to account of uncertainty in sample mean as well as to capture the possibility of greater sampling variability

  - Follows from 8.2 section

  - t-statistic has $(n-1)$ degrees of freedom

Definition 9.8. Test Statistic with Unknown Population Moment

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

Skytrax Airport Ratings Example: A business travel magazine wants to classify transatlantic gateway airports according to the mean rating for the population of business travelers. A rating scale with a low score of 0 and a high score of 10 will be used, and airports with a population mean rating greater than 7 will be designated as superior service airports

- Airports with 7 or more would be considered superior service airports
- Delhi Airport Authority wanted to test its service by conducting its own survey on these parameters.
    - Collected a sample of $n = 60$ passengers
    - $\bar{x} = 7.25$ and sample standard deviation was $s = 1.052$
- Is there enough evidence for Delhi to consider itself prepared for the next Skytrax survey? We conduct a hypothesis test based on this information but since we have no idea of what the population standard deviation looks like, we use the students-t distribution.

- Step 1: Developing a null hypothesis: rejection of null should lead to conclude that Delhi Airport has a ratings of more than 7 i.e. $H_a > 7$, an upper tailed test

$$H_0 : \mu \leq 7 \qquad H_a : \mu > 7$$

- Step 2: Decide the level of significance, $\alpha = 5\%$

- Step 3: Constructing a Test Statistic:

$$t = \frac{7.25 - 7}{1.052/\sqrt{60}} = 1.84$$

  - $(n - 1) = 59$ degree of freedom
  - p-value is $P(t \geq 1.84)$, since t-test is upper tail test

TABLE 2 *t* DISTRIBUTION (*Continued*)

Degrees of Freedom	Area in Upper Tail					
	.20	**.10**	**.05**	**.025**	**.01**	**.005**
35	.852	1.306	1.690	2.030	2.438	2.724
36	.852	1.306	1.688	2.028	2.434	2.719
37	.851	1.305	1.687	2.026	2.431	2.715
38	.851	1.304	1.686	2.024	2.429	2.712
39	.851	1.304	1.685	2.023	2.426	2.708
40	.851	1.303	1.684	2.021	2.423	2.704
41	.850	1.303	1.683	2.020	2.421	2.701
42	.850	1.302	1.682	2.018	2.418	2.698
43	.850	1.302	1.681	2.017	2.416	2.695
44	.850	1.301	1.680	2.015	2.414	2.692
45	.850	1.301	1.679	2.014	2.412	2.690
46	.850	1.300	1.679	2.013	2.410	2.687
47	.849	1.300	1.678	2.012	2.408	2.685
48	.849	1.299	1.677	2.011	2.407	2.682
49	.849	1.299	1.677	2.010	2.405	2.680
50	.849	1.299	1.676	2.009	2.403	2.678
51	.849	1.298	1.675	2.008	2.402	2.676
52	.849	1.298	1.675	2.007	2.400	2.674
53	.848	1.298	1.674	2.006	2.399	2.672
54	.848	1.297	1.674	2.005	2.397	2.670
55	.848	1.297	1.673	2.004	2.396	2.668
56	.848	1.297	1.673	2.003	2.395	2.667
57	.848	1.297	1.672	2.002	2.394	2.665
58	.848	1.296	1.672	2.002	2.392	2.663
59	.848	1.296	1.671	2.001	2.391	2.662
60	.848	1.296	1.671	2.000	2.390	2.660

- Step 3: Constructing a Test Statistic: for 59 *df*
    - From table we know p-value (p^*) for our t = 1.84 lies between 0.05 and 0.025
    - or we can surely say $p^* < \alpha$

- Step 4: Rejection using the p-Value: implies that when $p^* < \alpha$, then reject the null with at least 5% significance
    - Exact p-value i.e $P(t > 1.84)$ is $0.0354 < \alpha$: we reject the null hypothesis with 3% probability of making false rejection (Type I error) and claim that Delhi airport has good service quality not just based on our sample, but for the entire population

- Step 4B: Examine the Critical Value: of the test statistic
  - Given $\alpha = 5\%$, you need the upper area of t-statistic to be 0.05, given $df = 59$ or $t \geq 1.67$
  - Rejection for an upper tailed t-statistic implies: $t^* > t$, where t^* is calculated from the sample

- Hypothesis Result: We reject the null hypothesis that the population ratings is less than 7 with a 3% chance of making mistakes or with 97% (1-0.0354) confidence

9.4.1 Two Tailed Tests

Population Mean with Unknown σ

Holiday Toys Example distributes through 1000 retail outlets

- Wants to release a new toy before the Diwali season
- How much to produce?
- Expects 40 units per retail outlet, how much production?

Decides on a market survey to judge the demand, based on its expected toy features,

- Obtain a sample of 25 retailers, where retailer was asked to anticipate order quantity

Let μ denote the population mean order from each retailer.

Step 1: Developing a null hypothesis: H_0 cannot be rejected, then continue production as expected

$$H_0 : \mu = 40 \qquad H_a : \mu \neq 40$$

- Two tailed: because it wants to know in which direction is the production adjustment required if rejecting null.

Step 2: Decide the level of significance, α, say 5% error of rejecting null when its true or Type I error

Holiday Toys: sample collected gives $\bar{x} = 37.4$ and $s = 11.79$

Step 3: Constructing a test statistic, t using $\bar{x}, s$ from sample:

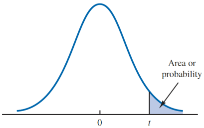
$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{37.4 - 40}{11.79/\sqrt{25}} = -1.10$$

Step 4: Compare p-value with α and reject/fail to reject null

- Since n = 25 and df = 24, check whether the sample is skewed or not.
- Assume it doesnt look skewed

▷ Step 4: Compare p-value with α and reject/fail to reject null

TABLE 2 t DISTRIBUTION



Entries in the table give t values for an area or probability in the upper tail of the t distribution. For example, with 10 degrees of freedom and a .05 area in the upper tail, $t_{.05} = 1.812$.

Degrees of Freedom	Area in Upper Tail					
	.20	.10	.05	.025	.01	.005
1	1.376	3.078	6.314	12.706	31.821	63.656
2	1.061	1.886	2.920	4.303	6.965	9.925
3	.978	1.638	2.353	3.182	4.541	5.841
4	.941	1.533	2.132	2.776	3.747	4.604
5	.920	1.476	2.015	2.571	3.365	4.032
6	.906	1.440	1.943	2.447	3.143	3.707
7	.896	1.415	1.895	2.365	2.998	3.499
8	.889	1.397	1.860	2.306	2.896	3.355
9	.883	1.383	1.833	2.262	2.821	3.250
10	.879	1.372	1.812	2.228	2.764	3.169
11	.876	1.363	1.796	2.201	2.718	3.106
12	.873	1.356	1.782	2.179	2.681	3.055
13	.870	1.350	1.771	2.160	2.650	3.012
14	.868	1.345	1.761	2.145	2.624	2.977
15	.866	1.341	1.753	2.131	2.602	2.947
16	.865	1.337	1.746	2.120	2.583	2.921
17	.863	1.333	1.740	2.110	2.567	2.898
18	.862	1.330	1.734	2.101	2.552	2.878
19	.861	1.328	1.729	2.093	2.539	2.861
20	.860	1.325	1.725	2.086	2.528	2.845
21	.859	1.323	1.721	2.080	2.518	2.831
22	.858	1.321	1.717	2.074	2.508	2.819
23	.858	1.319	1.714	2.069	2.500	2.807
24	.857	1.318	1.711	2.064	2.492	2.797

- Step 4: Compare p-value with α and reject/fail to reject null
  - Two-tailed test: so rejection region should be twice the area so computed.
  - From -1.10, we understand that area to the right would be between 20% and 10% probable i.e. between a t-value of 0.857 to 1.318
  - Since $p^*_{min} = 0.10 > \alpha = 0.05$, therefore H_0 cant be rejected with only 5% significance.
  - Computer generated p-value for t = -1.10 is 0.2822
  - Since $p^* = 0.2822 > \alpha = 0.05$, therefore H_0 cant be rejected with only 5% significance.

- Inference: There is not enough evidence that Holiday Toys Inc. should deviate from its original production plan

- Step 5: Compare critical value for the t-statistic
    - Under 24 df and $t_{.025} = 2.064$ and similarly $-t_{0.025} = -2.064$
    - Rejection rule is: to reject H_0 if $t^* \leq -2.064$ or $t^* \geq 2.064$
    - Inference: based on calculated t^*, we cannot reject the null. Therefore Holiday Toys should continue its production plans.

TABLE 9.3 SUMMARY OF HYPOTHESIS TESTS ABOUT A POPULATION MEAN: σ UNKNOWN CASE

	Lower Tail Test	Upper Tail Test	Two-Tailed Test
Hypotheses	$H_0: \mu \geq \mu_0$ $H_a: \mu < \mu_0$	$H_0: \mu \leq \mu_0$ $H_a: \mu > \mu_0$	$H_0: \mu = \mu_0$ $H_a: \mu \neq \mu_0$
Test Statistic	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$
Rejection Rule: *p*-Value Approach	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$
Rejection Rule: Critical Value Approach	Reject H_0 if $t \leq -t_\alpha$	Reject H_0 if $t \geq t_\alpha$	Reject H_0 if $t \leq -t_{\alpha/2}$ or if $t \geq t_{\alpha/2}$

9.5 Population Proportion

Denoting proportion's hypothesized value as p_0, we can have the following types of null and alternative

$$\overbrace{H_0 : p \geq p_0}^{\text{Lower Tailed Tests}} \qquad \overbrace{H_0 : p \leq p_0}^{\text{Upper Tailed Tests}} \qquad \overbrace{H_0 : p = p_0}^{\text{Two Tailed Tests}}$$

$$H_a : p < p_0 \qquad H_a : p > p_0 \qquad H_a : p \neq p_0$$

Hypothesis Testing of Proportions: methods are similar to sample mean (both in known and unknown σ case)

Difference: use sample proportions and its standard error to compute test statistic, *steps identical*.

- After choosing the significance level
- Construct the test statistic
- Use p-value or critical value of test static to reject/fail to reject the null

Population Proportion: An Example

The Bharatiya Pratham Party (BPP): is trying to promote female politicians to be in their central committee as an anti-discrimination representation effort

- Earlier years had only 20% women occupancy in their central committee
- Will the new efforts help them increase the proportions to a statistically significantly different level?

Step 1 involves formulating a null, H_0: to judge whether proportion has increased from 20%.

- Appropriate null: is $H_0 : p \leq 0.20$, a upper tail hypothesis – we are looking at rejecting the null, with some % Type I error.
- Appropriate alternative: is that the proportion of women politicians for BPP have increased i.e. $H_a : p > 0.20$

Step 2: Deciding a level of significance, α

- Party Manager (Mr Shodi): has asked you to keep a 5% significance level while testing their claim

- If H_0 can be rejected, the test results will give statistical support for the conclusion that the proportion of women increased and the campaign was beneficial.

Step 3: Construct a test statistic from a sample collected.

- Sample was collected by you and observed that 100 members were women from a 400 person sample all over India, i.e. $\bar{p}$ is the point estimator of p, which is unobserved.

- When H_0 is true $E(\bar{p}) = p_0$, where p_0 is the hypothesized value, with a standard error given as:

$$\sigma_{\bar{p}} = \sqrt{\frac{p_0(1-p_0)}{n}}$$

- Recall: From chapter 7 that when $np \geq 5$ and $n(1-p) \geq 5$, then sampling distribution of $\bar{p}$ can be represented by normal distribution.

- The test statistic then is:

$$z^* = \frac{\bar{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = 2.50$$

- Now you either need the p-value or you need the critical value of z-statistic to accept or reject the null

Step 4: Compute the p-Value

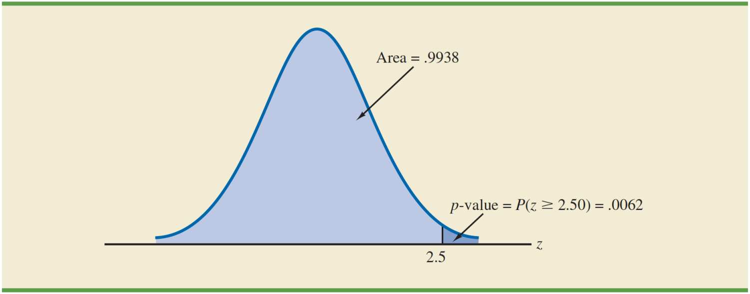
- Look up on the z-table for z = 2.50 implies a probability to the left of $z = 2.50$ to be a value of 0.9938.

- For an upper tail test, we need $P(z \geq 2.50) = 1 - P(z \leq 2.50) = 0.0062 = p^*$

- Rejection implies that the calculated p-value, p^* is less than the chosen significance level, α

Step 4: Compute the p-Value and rejection implies that the calculated p-value, p^* is less than the chosen significance level, α

- Since $p^* = 0.0062 < \alpha = 0.05$, we reject the null and provide evidence that their campaign increased the proportion of women in central committee, based on the sample.



Step 4A: Compute the critical-value: for a test statistic, $z_{0.05} = 1.645$

- Rejection rule: of H_0 when calculated z^* is greater than significant $z_{0.05}$, so $z^* = 2.50 > z_{0.05} = 1.645$

- p-value vs critical value: lead to the same hypothesis testing conclusion, but the p-value approach provides more information.

Population Proportion: Summary

TABLE 9.4 SUMMARY OF HYPOTHESIS TESTS ABOUT A POPULATION PROPORTION

	Lower Tail Test	Upper Tail Test	Two-Tailed Test
Hypotheses	$H_0: p \geq p_0$ $H_a: p < p_0$	$H_0: p \leq p_0$ $H_a: p > p_0$	$H_0: p = p_0$ $H_a: p \neq p_0$
Test Statistic	$z = \dfrac{\bar{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$	$z = \dfrac{\bar{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$	$z = \dfrac{\bar{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$
Rejection Rule: *p*-Value Approach	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$
Rejection Rule: Critical Value Approach	Reject H_0 if $z \leq -z_\alpha$	Reject H_0 if $z \geq z_\alpha$	Reject H_0 if $z \leq -z_{\alpha/2}$ or if $z \geq z_{\alpha/2}$

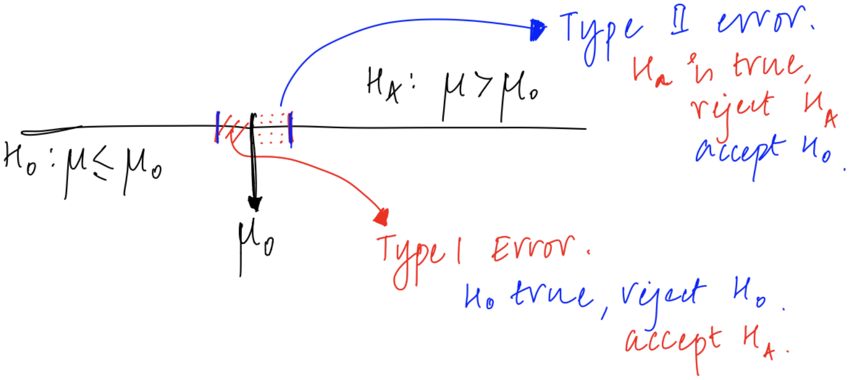
9.6 Hypothesis Testing and Decision Making

Important Inference Caveat:

- We choose a level of significance to determine rejection regions for both p-values or critical values of test statistic
- We only control for Type-I error since we assume H_0 is true when we reject it with α level of significance
- We do not control for Type-II error i.e. fail to reject null when null is false

Terminology: we recommend that we "do not reject the null" rather than stating that we "accept the null"

Summarizing errors and error regions under a one sided hypothesis



Decision making: requires not only accounting for Type-I error but also Type-II error, cause you want to bother about the total error

- When this occurs, decision making can be undertaken i.e. Null can be accepted or rejected with certainty

Lot acceptance sampling example:

- Decision: Quality manager must decide to accept a shipment from a supplier or return shippment due to poor quality
- Quality requirement: is 120 hours of battery life
- Sample tested: 36 batteries

Hypothesis formulated is:

$$H_0 : \mu \geq 120 \qquad H_a : \mu < 120$$

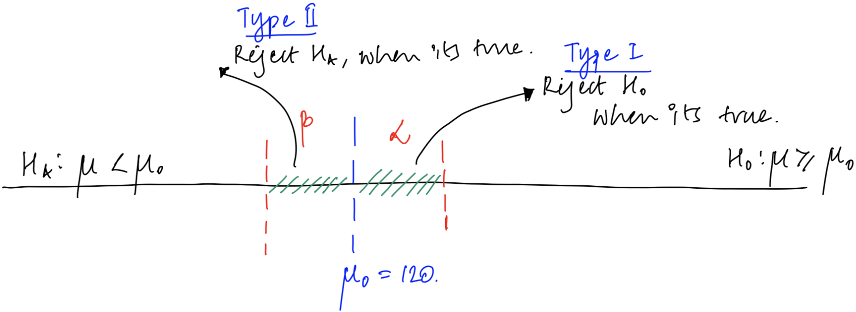
Resulting Decision:H_0 rejected implies that appropriate action is to return shipment. This includes region of Type I error plus H_a region.

- H_0 not rejected now includes region of Type II error.
- Section 9.7 p-value with Type II errors
- Section 9.8 adjustment of sample size to accommodate both Type I and Type II error commissions

9.7 Calculating the Probability of Type II Errors

- Lot acceptance sample: mean life of batteries in the shipment based on the sample of 36 batteries.

Summarizing Errors for Lot Acceptance Sample



Step 1: Calculate the rejection region for sample average

For what value of the sample mean can I reject the shipment based on the chosen level, α?

- $\alpha = 0.5$, $n = 36$, when $\sigma = 12$ the test statistic, z can be calculated for any $\bar{x}$

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{x} - 120}{12/\sqrt{36}}$$

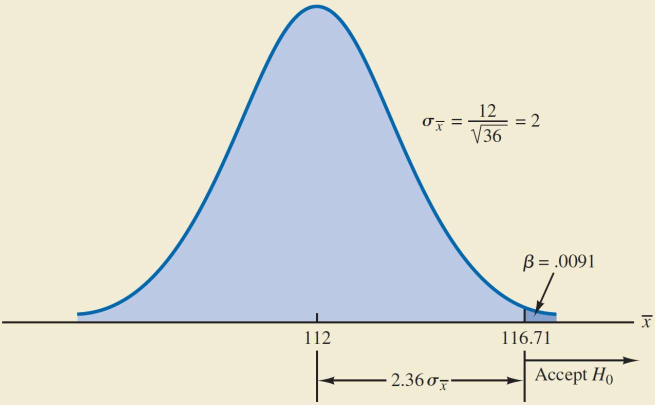
- Critical value of $z_{0.5} = 1.645$ with a lower tail rejection rule which implies if $z \leq -1.645$
- Solving for $\bar{x}$ under this rejection rule gives us $\bar{x} \leq 116.71$
- Decision rule: Accept the shipment if the sample mean $\bar{x} > 116.71$

Step II: Calculating the Type II error: is when we reject H_a when its true

- Actual average is less than $\mu < 120$ hours but we claim the average hours to be more than 120.
- For an assumed average μ (for instance $\mu = 112$), which is less than 120, we want to know what is the probability of accepting H_0
- This implies that we got a sample with mean, $\bar{x} \geq 116.71$, obtained from significance tests.
- What is the probability of sample mean greater than one obtained previously, $P(\bar{x} > 116.71)$

Assumed Population mean lower than choosen null, for Type II errors

FIGURE 9.8 PROBABILITY OF A TYPE II ERROR WHEN $\mu = 112$



Test statistic under counterfactual

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{116.71 - 112}{12/\sqrt{36}} = 2.36$$

Probability of z higher than 2.36 implies $P(1 - 0.9909) = 0.0091$

- There is less 0.9 percent probability of committing a Type II mistake when the actual $\mu = 112$, but we fail to reject the null

What if we the population mean μ is different from 112?

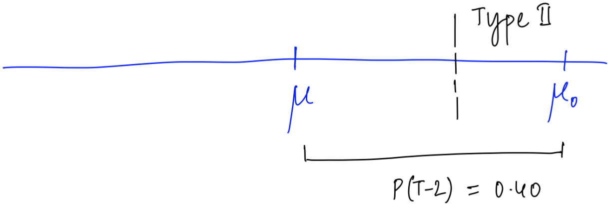
- Solve it for $\mu_a = 116$ to obtain 0.35 and the associated probability is 0.6368 or the probability to obtain a value of 116.71 is 0.3632.

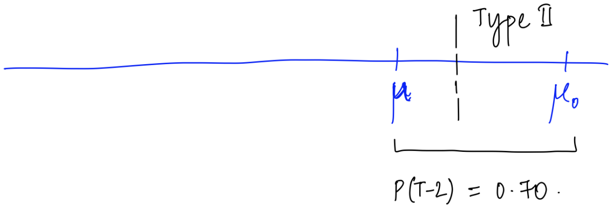
TABLE 9.5 PROBABILITY OF MAKING A TYPE II ERROR FOR THE LOT-ACCEPTANCE HYPOTHESIS TEST

Value of μ	$z = \dfrac{116.71 - \mu}{12/\sqrt{36}}$	Probability of a Type II Error (β)	Power ($1 - \beta$)
112	2.36	.0091	.9909
114	1.36	.0869	.9131
115	.86	.1949	.8051
116.71	.00	.5000	.5000
117	−.15	.5596	.4404
118	−.65	.7422	.2578
119.999	−1.645	.9500	.0500

Figure 2: when the true population mean μ is close to the null hypothesis value of $\mu = 120$, the probability is high that we will make a Type II error.

As actual μ nears assumed μ_0 under null, probability of making Type II error increases





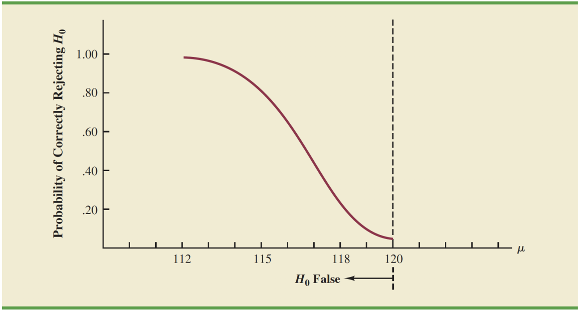
Power of Test: the probability of correctly rejecting H_0 when it is false.

- For a value of μ, power is $(1-\beta)$

Power curve extends over the values of μ for which the null hypothesis is false.

- Height of the power curve at any value of μ indicates the probability of correctly rejecting H_0 when it is false

FIGURE 9.9 POWER CURVE FOR THE LOT-ACCEPTANCE HYPOTHESIS TEST



9.8 Determining the Sample Size

Level of significance: gives you a method of controlling the possible Type 1 error that we commit

Sample size: gives you control over Type II error too! Consider the null and the alternative as follows:

$$H_0 : \mu \geq \mu_0 \qquad H_a : \mu < \mu_0$$

- Step 1: Get an $\bar{x}$ value for which we commit α percent Type I error
- Step 2: Get an $\bar{x}$ value for which we commit β percent Type II error
- Step 3: Equate the $\bar{x}$ in both conditions and solve for the optimal size of the sample

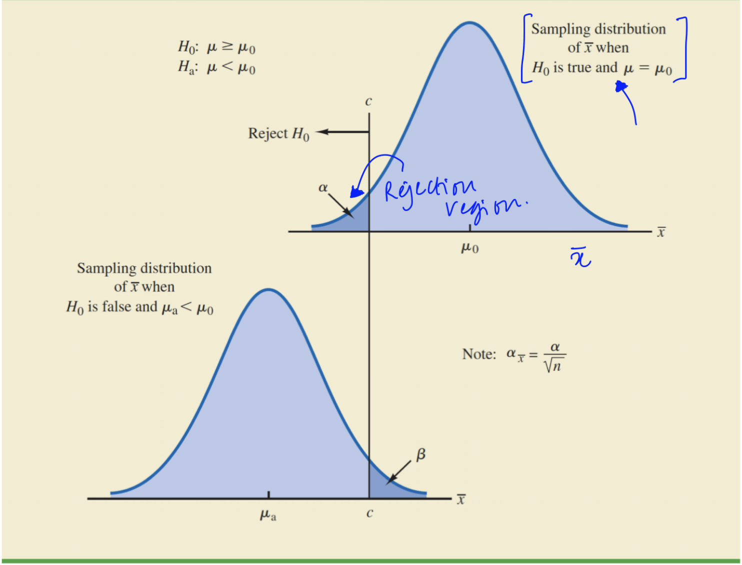
Step 1: Get an $\bar{x}$ value for which we commit α percent Type I error (upper panel)

- Lower tailed test: the critical value $-z_\alpha$ test statistic
- Given σ (whether known or unknown), a sample size n, and hypothesized value of population mean μ_0
- Rejection region is when $\bar{x} \leq c$ and the $\mathrm{P}(\bar{x} \leq c) = \alpha$

$$c = \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}}$$

c is that value of '$\bar{x}$' which you obtained when solving for it using critical value approach as in the Lot acceptance example (c = 116.71)

FIGURE 9.10 DETERMINING THE SAMPLE SIZE FOR SPECIFIED LEVELS OF THE TYPE I (α) AND TYPE II (β) ERRORS

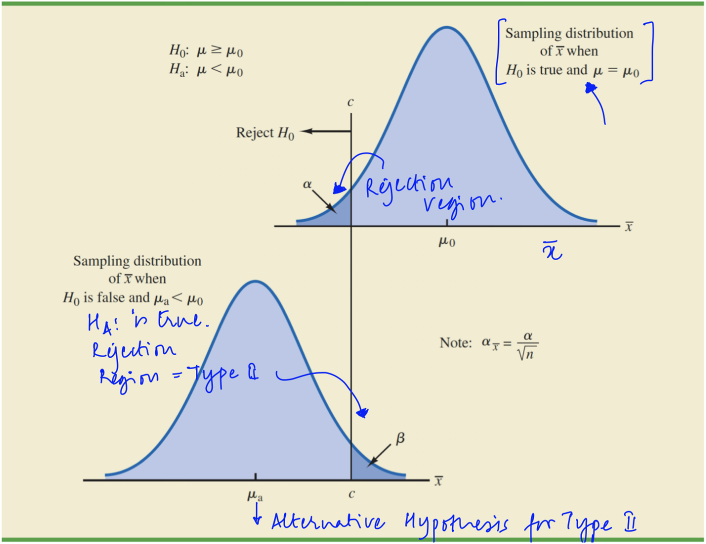


Step 2: Get an $\bar{x}$ value for which we commit β percent Type II error (lower panel)

- Sampling distribution of $\bar{x}$ when alternative hypothesis is true i.e. $\mu = \mu_a < \mu_0$
- β region shows the Type II error as calculated for $\mu_a = 112$ in the Lot acceptance example

$$c = \mu_a - z_\beta \frac{\sigma}{\sqrt{n}}$$

FIGURE 9.10 DETERMINING THE SAMPLE SIZE FOR SPECIFIED LEVELS OF THE TYPE I (α) AND TYPE II (β) ERRORS



Step 3: Equate the $\bar{x}$ in both conditions and solve for the optimal size of the sample

- Equate the c, so that the probability of Type I error is α and of Type II error is β

SAMPLE SIZE FOR A ONE-TAILED HYPOTHESIS TEST ABOUT A POPULATION MEAN

$$n = \frac{(z_\alpha + z_\beta)^2\sigma^2}{(\mu_0 - \mu_a)^2} \tag{9.7}$$

where

z_α = z value providing an area of α in the upper tail of a standard normal distribution
z_β = z value providing an area of β in the upper tail of a standard normal distribution
σ = the population standard deviation
μ_0 = the value of the population mean in the null hypothesis
μ_a = the value of the population mean used for the Type II error

Note: In a two-tailed hypothesis test, use (9.7) with $z_{\alpha/2}$ replacing z_α.

Lot Acceptance example: They allow an α of 0.05 and a β of 0.10. Determine the sample size?

$$n = \frac{(1.645 - 1.28)^2 12^2}{(120 - 115)^2} = 49.3 \simeq 50$$

- Managers can make accept/reject H_0 statements now since they have both Type I and Type II errors under consideration

Relation between α, β and sample size n

1. We determine α, β and n jointly: we know two, can obtain the third.

2. Given α, a larger n will reduce $\beta \rightarrow$ large sample always better

3. Given n, decreasing α will increase β

Important: If you chooses small values of α, then you increase your chance of Type II error. Choosing too small a significance level might penalize you.