Lec 6*: Hypothesis Testing

1 Motivation

- After all the prep work we have done throughout the past couple lectures, we are now finally ready to utilize the power of **inferential statistics**.
- An outline of how we are going to make inference about the population:
  - We first hypothesize a test about the population parameter.
  - Then, using the known sample statistic, and the distribution that such sample statistic follows, we check whether the realized value of our sample statistic would support the hypothesis.
  - If so, our default hypothesis seems to be our best shot.
  - If not, at least we know that we cannot accept the default hypothesis.

2 Hypothesis

- When performing hypothesis testing, you can divide the outcome of your test into two hypotheses:
  1. The default outcome called **the null hypothesis**, denoted as H_0
  2. The alternative outcome called **the alternative hypothesis**, denoted as H_1
- How do we know which outcome is the null, and which one is the alternative?
  - By convention, the null hypothesis is where you write the outcome with equality.
  - For the outcome that doesn't have an equality sign (i.e. the outcome with $\neq$, $>$, or $<$ sign), it goes to the alternative hypothesis.

Exercise. Write down the null and alternative hypothesis for each proposed testing scenario:

1. Whether μ equals 5.
   $H_0 : \mu = 5$
   $H_1 : \mu \neq 5$
2. Whether μ is greater than 10.
   $H_0 : \mu = 10$
   $H_1 : \mu > 10$
3. If μ is less than 15.
   $H_0 : \mu = 15$
   $H_1 : \mu < 15$

- What do we do once we have the null and alternative hypothesis?
  - Test **whether we can reject the null**
    - If we **reject the null**, then we **accept the alternative** (under conditions that lead to the rejection).

*Some exercise questions are taken from or slightly modified based on Dr. Gregory Pac's Econ 310 discussion handout.

* If we **fail to reject the null**, then it **doesn't necessarily mean that the null is true**. It just happens to be that, compared to the given alternative hypothesis, the null is the more likely outcome.

3 Performing a Hypothesis Test

- So how do we test whether we can reject the null?
- In this section, we are going to use the example of testing null hypothesis related to the population mean μ, where the corresponding sample statistic is $\bar{X}$.
- To see whether the $\bar{X}$ we get from a sample supports the rejection or failure of rejection of the null, it would be very helpful to know how $\bar{X}$ is distributed; this way, we would know how $\bar{X}$ compares to the proposed population mean level from the null hypothesis.

3.1 What is the distribution of $\bar{X}$?

- From Lec 6, we learned about the sampling distribution of $\bar{X}$:
  1. If X is exactly normally distributed, then $\bar{X}$ is also exactly normally distributed.
  2. If X is not exactly normally distributed, but the sample size $n \geq 30$, then we can apply central limit theorem (CLT), and claim that $\bar{X}$ is approximately normally distributed.
  3. If X is not exactly normally distributed, and the sample size $n < 30$, then we're screwed.
- So in order to perform hypothesis testing, we need to be in case 1 or 2, and **we are going to assume for the rest of this handout that either case 1 or 2 hold**.
- Under case 1 or 2, we know that a normal distribution is needed to describe the relationship between $\bar{X}$ and the hypothesized μ.
- Now begs the question: if $\bar{X}$ follows a normal distribution (either exactly or approximately), what are the parameters of this normal distribution?
  - The variance of the normal distribution is pretty straight forward, since we are still assuming that the population standard deviation of X is known as σ, and the sample size n is given to you. Thus,

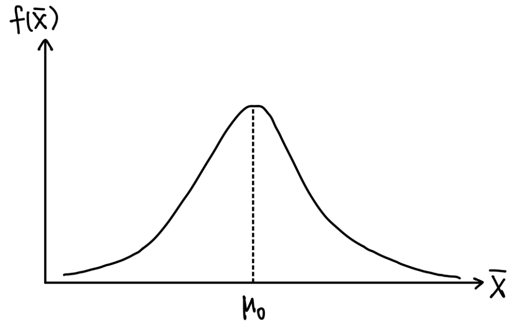
    $$\text{Variance of the normal distribution} = \frac{\sigma^2}{n} = \left(\frac{\sigma}{\sqrt{n}}\right)^2$$

    (Implicitly assuming that N is large enough so that finite population correction factor for the standard deviation of sample mean is not needed.)
  - What about the mean of the normal distribution then? This is the item that is unknown, but we have some hypothesized value about it. Specifically, the null hypothesis assumes that the mean of the normal distribution equals to a certain value. We are going to call the value of population mean given in H_0 as μ_0.
    For example, if $H_0 : \mu = 4$, then $\mu_0 = 4$. Anyways,

    $$\text{Mean of the normal distribution} = \mu_0$$

- Thus, $\bar{X}$ follows (exactly or approximately)

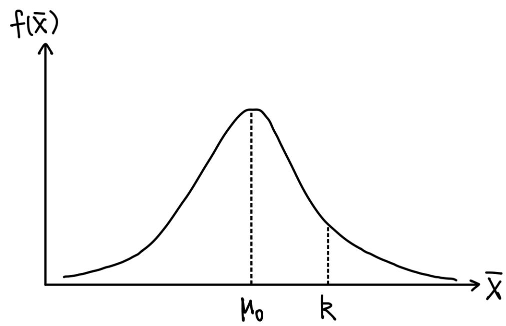
$$N\left(\mu_0, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$$



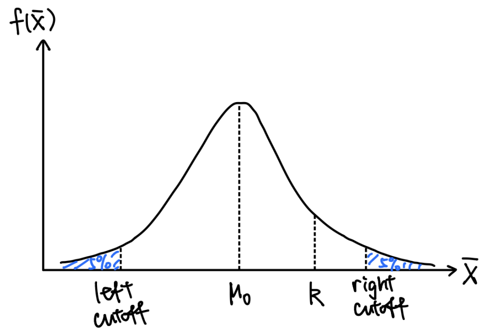
3.2 Test method 1: test statistic and rejection region

- Say that the sample mean calculated is $k > 0$, and our hypotheses are the following:

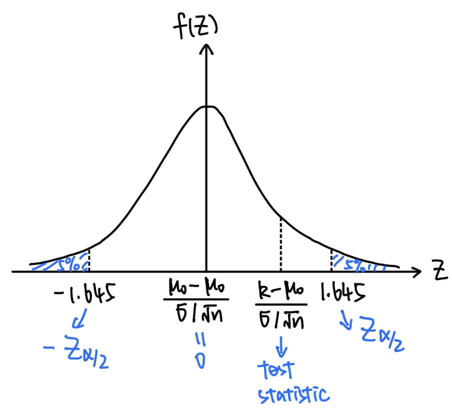
$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$



- To reject the null and accept the alternative, we need k to be **as far away from the hypothesized mean μ_0 as possible** – either to on the far left side of μ_0, or on the far right side of μ_0.
  - Exactly how far away from μ_0 is a choice made by the researcher, and this is where **significance level** (α) comes into play.
- Say that one sets the significance level $\alpha = 10\%$. Then we'd like the probability of being in the far left side and the far right side of the distribution to sum up to be 10% – pretty unlikely outcome, so if we reach these far sides, we know that the null of $\mu = \mu_0$ is unlikely to hold.
  - Since the far left and far right side probability sum up to be 10%, people tend to just equally divide them for each tail:



- So if $k <$ left cutoff, or if $k >$ right cutoff, then we reject the null at significance level $\alpha = 10\%$.
- There is just one drawback to this whole plan so far: $\bar{X}$ doesn't follow a standard normal distribution, so the left and right cutoff value will change all the time. To make it easier, we can standardize $\bar{X}$ so that a standard normal distribution is followed. This means that

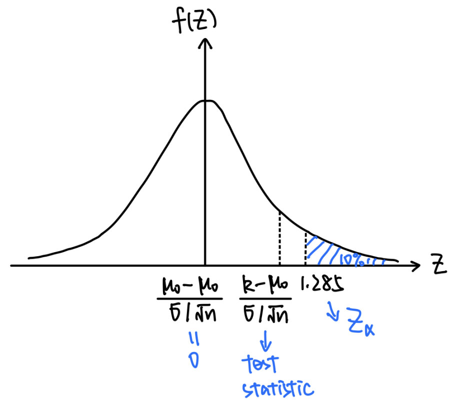


- Now, if $\frac{k-\mu_0}{\sigma/\sqrt{n}}$ is less than $-Z_{\alpha/2} = -1.645$, or if $\frac{k-\mu_0}{\sigma/\sqrt{n}}$ is greater than $Z_{\alpha/2} = 1.645$, then we reject the null at $\alpha = 10\%$ confidence level.
  - **Test statistic** $= \frac{k-\mu_0}{\sigma/\sqrt{n}}$ (standardized sample mean)
  - **Rejection region**: $|\text{test statistic}| > Z_{\alpha/2}$

The aforementioned method considers an alternative hypothesis $H_1 : \mu \neq \mu_0$, which is why we reject when test statistic is either too far to the left, or too far to the right of the standardized sample mean distribution. Such test is called a **two-tailed test**.

If H_1 only considers one side (i.e. a **left-tailed / right-tailed test**; e.g. $H_1 : \mu > \mu_0$ is a right-tailed test), then we only need to see if the test statistic is too far to one side of the sample mean distribution.

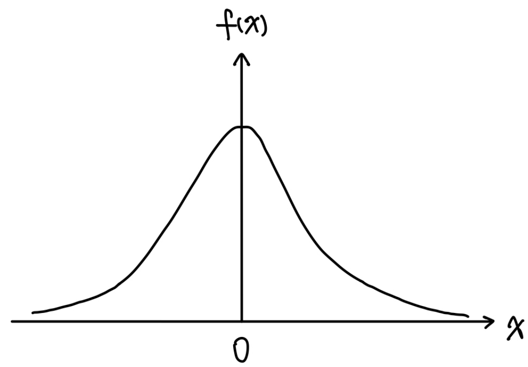
If we continue to consider the sample mean calculated as k, the significance level chosen as 10%, and a right-tailed test is performed, then



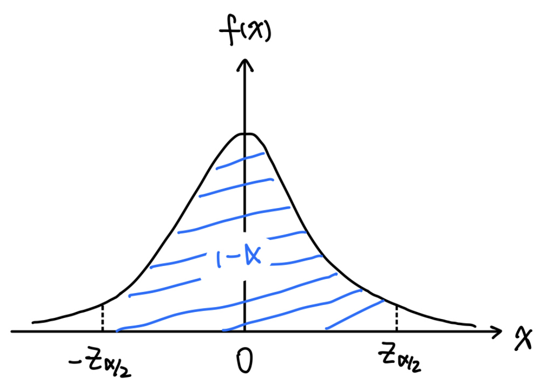
Rejection region: test statistic $= \frac{k-\mu_0}{\sigma/\sqrt{n}} > Z_\alpha$

3.3 Test method 2: confidence interval test (for a two-tailed test only)

- First thing first, what is a confidence interval?
  - Think about a standard normal distribution:



  - If we want to cover $(1-\alpha)$ portion of this standard normal distribution, then



  - We can think about this standard normal distribution as the sampling distribution of the mean, where the sample mean estimator has been standardized:

$$P\left(-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}\right) \qquad = 1-\alpha$$

$$
\begin{aligned}
\Leftrightarrow \quad & P\left(-Z_{\alpha/2} \leq \frac{\bar{X}-\mu_X}{\sigma_X/\sqrt{n}} \leq Z_{\alpha/2}\right) && = 1-\alpha \\
\Leftrightarrow \quad & P\left(-Z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}} \leq \bar{X}-\mu_X \leq Z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}}\right) && = 1-\alpha \\
\Leftrightarrow \quad & P\left(\bar{X}-Z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}} \leq \mu_X \leq \bar{X}+Z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}}\right) && = 1-\alpha
\end{aligned}
$$

Thus, for $(1-\alpha)$ portion of area covered, the confidence interval constructed is

$$\left[\bar{X}-Z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}}, \bar{X}+Z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}}\right]$$

We call $(1-\alpha)$ the **confidence level** for the above interval.

– What are some common confidence level and the associated Z score ($Z_{\alpha/2}$)?

Confidence level	α	$Z_{\alpha/2}$
90%	0.1	1.645
95%	0.05	1.96
99%	0.01	2.575

– **Interpretation**:

Say that, for example, a 95% confidence interval of the mean of X, using a sample of size 70, is estimated to be $[4, 8]$. The following are some examples of correct interpretation of this confidence interval constructed.

* **Correct version 1**: There's a 5% probability that the population mean of X lies outside of the confidence interval estimator. For this sample of size 70, we estimate the confidence interval to be $[4, 8]$.
* **Correct version 2**: If random sample of size 70 were repeatedly selected, then in the long run, 95% of the confidence intervals formed would contain the true mean of X, which in this case is between 4 and 8.

Exercise. Suppose you draw a sample from a population with a standard deviation of 25. You draw 50 observations and end up with a sample mean of 100.

Estimate a 90% confidence interval for the population mean

The confidence interval estimator is the following:

$$
\begin{aligned}
\left[\bar{X}-Z_{0.10/2}\frac{\sigma_X}{\sqrt{n}}, \bar{X}+Z_{0.10/2}\frac{\sigma_X}{\sqrt{n}}\right] &= \left[100-1.645\times\frac{25}{\sqrt{50}}, 100+1.645\times\frac{25}{\sqrt{50}}\right] \\
&= [94.18, 105.82]
\end{aligned}
$$

Thus, the 90% confidence interval is estimated to be $[94.18, 105.82]$.

- Now, think about how confidence interval is constructed – it sounds awfully a lot like the reversed process compared with performing a two-tailed test using test statistic and rejection region:

- Confidence interval gives you the region to **not reject** the null under significance level α.

- That is, for

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

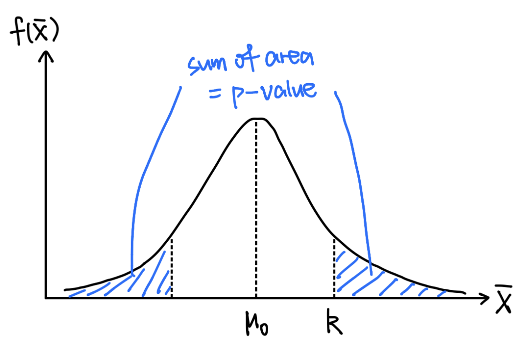
  - If $\mu_0 \in$ CI constructed at $(1-\alpha)$ confidence level, then we fail to reject the null at α significance level
  - If $\mu_0 \notin$ CI constructed at $(1-\alpha)$ confidence level, then we reject the null at α significance level

3.4 Test method 3: p-value

- Say that we are still performing a two-tailed test:

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

- Instead of looking at which region to reject given a significance level, we can think about if the current sample mean is used as the cutoff value, then what's the probability that we can reject the null hypothesis – this is the **p-value**:



$$\begin{aligned}\text{p-value} &= P(\bar{X} > k) \times 2 \\ &= P\left(\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} > \frac{k-\mu_0}{\sigma/\sqrt{n}}\right) \times 2 = P\left(Z > \frac{k-\mu_0}{\sigma/\sqrt{n}}\right) \times 2\end{aligned}$$

(Note: for example, if a right-tailed test is performed, then p-value $= P\left(Z > \frac{k-\mu_0}{\sigma/\sqrt{n}}\right)$)

- Now, since p-value represents the probability that we can reject the null given the sample mean k,
  - If p-value $>$ significance level α, we fail to reject the null
  - If p-value $<$ significance level α, we reject the null

  Essentially, **p-value is the smallest significance level needed to reject the null**.

Note:

1. If you have a two-tailed test, you can use any of the three methods.
2. If you have a one-tailed test, you can only use method 1 (test statistic and rejection region) or method 3 (p-value).
3. **No matter which test method is used, the same conclusion should be reached**

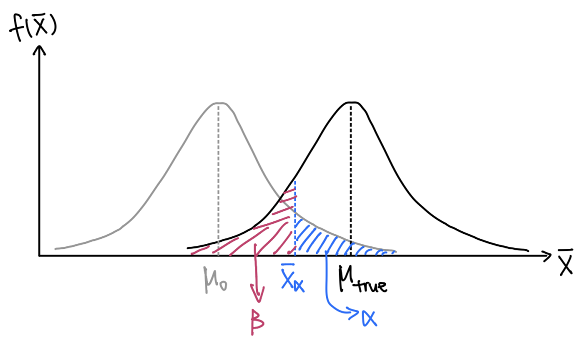
4 Errors When Performing a Hypothesis Test

	H_0 is True	H_0 is False
Reject H_0	**Type I Error** Probability $= \alpha$	**Power** Probability $= 1 - \beta$
Not Reject H_0	GOOD JOB Probability $= 1 - \alpha$	**Type II Error** Probability $= \beta$

- The way we construct the aforementioned tests is by setting the level of Type I Error (i.e. setting what α is), since we think rejecting the null when it is true is a lot more costly than not rejecting the null when it is false (recall the criminal justice analogy from lecture).
- Sometimes, if you have information about what the true population mean is, it is helpful to quantify the probability of committing Type II Error as well; this way, we can also evaluate how the tests perform when the null should be rejected.
- Say that we have a right-tailed test

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu > \mu_0$$

Let the true population mean be μ_{true}, and let $\bar{X}_\alpha$ be the point where significance level is set as α:



Thus, the probability of committing Type II error $= \beta = P(\bar{X} < \bar{X}_\alpha | \mu = \mu_{true}) = P(Z < \frac{\bar{X}_\alpha - \mu_{true}}{\sigma/\sqrt{n}})$

5 Exercises

1. Suppose a ski resort bases revenue projections on the assumption that an average skier skis four times per year. To evaluate the validity of this assumption, a random sample of 63 skiers is drawn and each

is asked to report the number of times he or she skied last year. The sample average is 4.38 times. Assuming the population standard deviation is 2 times, can we infer at a 10% significance level that the resort's assumption is wrong?

(a) Set the null and alternative hypotheses.

The null hypothesis is $H_0 : \mu = 4$.

The alternative hypothesis is $H_1 : \mu \neq 4$.

(b) Calculate a test statistic, select a rejection region, draw a conclusion, and interpret this conclusion.

Since $n = 63 \geq 30$, by central limit theorem, $\bar{X} \overset{a}{\sim} N\left(\mu_0, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$.

Given that this is a two-tailed test with $\alpha = .1$, the rejection region is $|\text{test statistic}| > 1.645$.

For this sample, the test statistic $= \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} = \frac{4.38-4}{2/\sqrt{63}} = 1.51$.

Since $|\text{test statistic}| > 1.645$ is NOT satisfied, we fail to reject the null at 10% significance level.

Interpretation: at a 10% significance level, we cannot rule out the resort's assumption that the average skier skis four times per year.

(c) Calculate a 90% confidence interval for the population mean. Repeat the test in part (b)using the confidence interval method.

The 90% confidence interval is

$$\begin{aligned} &\left[\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \\ &= \left[4.38 - 1.645 \times \frac{2}{\sqrt{63}}, 4.38 + 1.645 \times \frac{2}{\sqrt{63}}\right] \\ &= [3.97, 4.79] \end{aligned}$$

Since $\mu_0 = 4$ is inside the confidence interval, and the confidence interval is the complement of the rejection region, we fail to reject the null hypothesis. So the result and interpretation are identical as they were in part (b). More generally, the confidence interval method and the rejection region method always yield identical results.

(d) Repeat the test in part (b) using the p-value method.

To obtain the p-value, we want the probability of observing a test statistic at least as extreme as 1.51 (assuming the null is true). So we have

$$\begin{aligned} \text{p-value} &= 2 \times P(\bar{X} > 4.38) \\ &= 2 \times P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{4.38 - 4}{2/\sqrt{63}}\right) \\ &= 2 \times P(Z > 1.51) = 2 \times [1 - P(Z \leq 1.51)] = 0.1310 \end{aligned}$$

Since the p-value $= 0.1310 > 0.10 = \alpha$, we fail to reject the null hypothesis. So the result and interpretation are identical as they were in part (b). More generally, the p-value method and the rejection region method always yield identical results.

2. A light bulb company claims that, on average, its bulbs last more than 5,000 hours. A random sample of 100 bulbs is drawn and the sample mean bulb life is 5,120 hours. Assuming the population standard deviation is 400 hours, test the company's claim using a 5% significance level.

(a) Set the null and alternative hypotheses.

The null hypothesis is $H_0 : \mu = 5000$.

The alternative hypothesis is $H_1 : \mu > 5000$.

(b) Calculate the test statistic, select a rejection region, draw a conclusion, and interpret this conclusion.

Since $n = 100 \geq 30$, by central limit theorem, $\bar{X} \overset{a}{\sim} N\left(\mu_0, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$.

Given that this is a right-tailed test with $\alpha = .05$, the rejection region is test statistic > 1.645.

For this sample, the test statistic $= \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} = \frac{5120-5000}{400/\sqrt{100}} = 3$.

Since test statistic > 1.645 is satisfied, we reject the null at 5% significance level.

Interpretation: at a 5% significance level, we can reject the null hypothesis and conclude that light bulbs last, on average, more than 5,000 hours.

(c) Repeat the test in part (b) using the p-value method.

To obtain the p-value for a right-tailed test, we want:

$$\begin{aligned}
\text{p-value} &= P(\bar{X} > 5120) \\
&= P\left(\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} > \frac{5120-5000}{400/\sqrt{100}}\right) \\
&= P(Z > 3) = 1 - P(Z \leq 3) = 0.0013
\end{aligned}$$

Since the p-value $= 0.0013 < 0.05 = \alpha$, we reject the null hypothesis. So the result and interpretation are identical as they were in part (b). More generally, the p-value method and the rejection region method always yield identical results.

3. You wish to test the null hypothesis $H_0 : \mu = 200$ against the alternative that $H_1 : \mu < 200$. You draw a sample of size $n = 9$ and obtain a sample mean of 190. You may assume that the population is normally distributed with a standard deviation of 50.

(a) Compute and interpret the p-value of the test.

Since the population is normally distributed, $\bar{X} \sim N\left(\mu_0, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$.

This is a left-tailed test, so the p-value is:

$$\begin{aligned}
\text{p-value} &= P(\bar{X} < 190) \\
&= P\left(\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} < \frac{190-200}{50/\sqrt{9}}\right) \\
&= P(Z < -0.6) = P(Z > 0.6) = 1 - P(Z < 0.6) = 0.2743
\end{aligned}$$

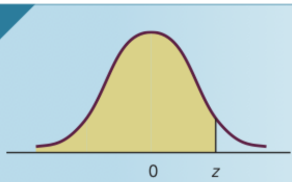
Interpretation: a significance level of $\alpha = 0.2743$ is the smallest significance level at which we can reject the null.

(b) Find the p-value if the standard deviation had been $\sigma = 30$ and $\sigma = 10$, respectively. Comment on your findings.

For $\sigma = 30$, the test statistic would be $z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} = \frac{190-200}{30/\sqrt{9}} = -1$, so the p-value would be $P(z < -1) = 0.1587$.

For $\sigma = 10$, the test statistic would be $z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} = \frac{190-200}{10/\sqrt{9}} = -3$, so the p-value would be $P(z < -3) = 0.0013$.
So holding $\bar{X}$, μ_0 and n constant, the p-value decreases as σ decreases. This means that as σ decreases, it becomes easier to reject the null hypothesis.

TABLE **3** ***(Continued)***



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

Z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990