Section 5.2: Hypothesis Tests for $\mu$

- Announcements:
  - Tutoring: Math office in 111 Cummington 1st floor, Rich Hall.
  - Away week of November 13. Xinyi will fill in.
  - Cheat sheet - 1 pg, single sided.
  - Office hours next Friday moved to next Wednesday so you guys can ask me questions before the exam.

3.2: Using $p$-values

- ($\sigma$ known)
  1. Assume that,
     (a) The sample is obtained using simple random sampling,
     (b) The sample has no outliers and the population from which the sample is drawn is normally distributed or that $n \geq 30$.
  2. Step 1: Determine if the test is a two, left, or right-tailed test.
  3. Select a level of significance $\alpha$ based on the seriousness of making a Type I error, and compute
     \[ P[Z > z_{\alpha}] = \alpha, \quad Z \sim N(0, 1) \]
     if left or right-tailed or $P[Z > z_{\alpha/2}]$ if two-tailed.
  4. Compute
     \[ z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \]
     which is the number of standard errors the sample mean $\bar{X}$ is from the null hypothesis mean $\mu_0$.
  5. If,
     (a) (two-tailed) $P[Z > |z_0|] < \alpha$ reject the null hypothesis.
     (b) (left-tailed) $P[Z < z_0] < \alpha$ reject the null hypothesis.
• These $p$-values quantify the probability that a sample will result in a sample mean such as the one obtained if the null hypothesis is true. Thus, we reject the null hypothesis if the $p$-value is too small (so that it’s very unlikely that we got the sample mean we did assuming the null hypothesis is true) and so yields the exact significance of the data.

• ($\sigma$ unknown)
  1. Assume that,
    (a) The sample is obtained using simple random sampling,
    (b) The sample has no outliers and the population from which the sample is drawn is normally distributed or $n \geq 30$.
  2. Step 1: Determine if the test is a two, left, or right-tailed test.
  3. Select a level of significance $\alpha$ based on the seriousness of making a Type I error, and compute
     \[ P[Z > t_{\alpha}] = \alpha, \quad Z \sim t_{dist}(n - 1) \]
     if left or right-tailed or $P[Z > z_\frac{\alpha}{2}]$ if two-tailed.
  4. Compute
     \[ t_0 = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \]
     is the number of sample standard errors away $\bar{X}$ is from $\mu_0$.
  5. If,
    (a) (two-tailed) $P[Z > |t_0|] < \alpha$ reject the null hypothesis.
    (b) (left-tailed) $P[Z < t_0] < \alpha$ reject the null hypothesis.
    (c) (right-tailed) $P[Z > t_0] < \alpha$ reject the null hypothesis.

• Examples:
  – An engineer wants to measure the bias in a pH meter. She uses the meter to measure the pH in 14 natural substances and obtains the following data,
    \[ 7.01, 7.04, 6.97, 7.00, 6.99, 6.97, 7.04, 7.04, 7.01, 7.00, 6.99, 7.04, 7.07, 6.97. \]
    Is there sufficient evidence to support the claim that the pH meter is not correctly calibrated at $\alpha = 0.05$?
    **Solution:**
    The data’s free of outliers and approximately normal, so we’re good.
    This is a two-tailed test.
    $\mu_0 = 7.0, \bar{X} = 7.01, s = 0.032, n = 14.$
    \[ t_0 = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{7.01 - 7.0}{0.032/\sqrt{14}} = 1.1693. \]
    So, $P[Z > |t_0|] = P[Z > t_0] + P[Z < -t_0] = 2 \cdot P[Z > t_0] = 2 \cdot P[Z > 1.1693] = 2 \cdot 0.1316 = 0.2633.$
Since $0.2633 > 0.05$, we cannot reject the null hypothesis, so that with a 5% level of significance, the pH meters are calibrated correctly.

In this example, we know the exact level of significance in the data, namely 35%. This is the smallest level of significance we need in order to still reject $H_0$.

### 3.3: Analysis and Conclusions of the Hypothesis Tests

- **(Power)** When performing hypothesis testing, we control $\alpha$ by prescribing it outright, where recall that,

$$\alpha = \mathbb{P} [\text{Type I Error}] = \mathbb{P} [\text{Reject } H_0|H_0 \text{ True}].$$

$$\beta = \mathbb{P} [\text{Type II Error}] = \mathbb{P} [\text{Accept } H_0|H_1 \text{ True}].$$

$\beta$ on the other hand is difficult to control since it depends on several factors.

- One of these parameters is $\alpha$ itself,

$$\beta = \mathbb{P} [H_0 \text{ True}|H_1 \text{ True}] = \mathbb{P} [H_1 \text{ True}|H_0 \text{ True}] \cdot \frac{\mathbb{P} [H_0 \text{ True}]}{\mathbb{P} [H_1 \text{ True}]}$$

$$= (1 - \alpha) \cdot \frac{\mathbb{P} [H_0 \text{ True}]}{\mathbb{P} [H_1 \text{ True}]}$$

Thus, increasing $\alpha$ decreases $\beta$, so that one must weight the choice of a smaller $\beta$, which is desirable, with a larger level of significance, $\alpha$, which is undesirable.

- The *power* of a test is defined to be the ability of the test to detect or reject a false null hypothesis. Quantitatively, it’s defined by,

$$\text{Power} = 1 - \beta = \mathbb{P} [\text{Reject } H_0|H_0 \text{ False}] = \mathbb{P} [\text{Reject } H_0|H_1 \text{ True}].$$

- As power increases, $\beta$ decreases, resulting in a better test (lower type II error).

- Power is a complicated function of three variables:
  1. $n=$Sample Size,
  2. $\alpha=$Level of Significance=$\mathbb{P} [\text{Type I Error}]$,
  3. $ES=$The Effect Size=The standardized difference in means specified under $H_0$ and $H_1$ defined by,

$$ES = \frac{|\mu_0 - \mu_1|}{\sigma}.$$ 

- As such, we’ll focus only on the case where $\sigma$ is known.

- Let’s investigate the effect of each of these parameters on power.
Figure 1: $\bar{X}$ under $H_0$ with mean $\mu_0 = 100$.

Figure 2: $\bar{X}$ under $H_0$ with mean $\mu_0 = 100$ and $H_1$ with $\mu_1 = 110$. 
Figure 3: $\bar{X}$ under $H_0$ and $H_1$ with $\alpha = 0.05$. 
Figure 4: Changing $\alpha$ to 0.1 fixing $\mu_1 = 110$.

Figure 5: Changing $\mu_1$ to 120 with fixing $\alpha = 0.05$. All pictures from D’Agostino et al.
– The last effect is changing the sample size \( n \), but we’ve done this before: a larger sample size ensures a more powerful test.

– Let’s compute the power \( 1 - \beta \) where, for now, we’re considering a right-tailed test. Recall that \( z_\alpha \) is defined by,

\[
P[Z > z_\alpha] = \alpha
\]

where \( Z = \frac{\bar{X}_\alpha - \mu_0}{\sigma/\sqrt{n}} \). Thus,

\[
P \left[ \frac{\bar{X}_0 - \mu_0}{\sigma/\sqrt{n}} > z_\alpha \right] = \alpha
\]

\[\Leftrightarrow P \left[ X_0 > \frac{\sigma}{\sqrt{n}} z_\alpha + \mu_0 \right] = \alpha\]

and so \( x_\alpha = \frac{\sigma}{\sqrt{n}} z_\alpha + \mu_0 \) in the above picture.

Now,

\[
\bar{X}_1 = \frac{\sigma}{\sqrt{n}} Z + \mu_1
\]

so that,

\[
1 - \beta = P \left[ \bar{X}_1 > x_\alpha \right] = P \left[ \frac{\sigma}{\sqrt{n}} Z + \mu_1 > \frac{\sigma}{\sqrt{n}} z_\alpha + \mu_0 \right]
\]

\[
\Rightarrow 1 - \beta = P \left[ Z > z_\alpha - \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} \right].
\]

Since this is a right-tailed test, \( \mu_1 > \mu_0 \) so that

\[
1 - \beta = P \left[ Z > z_\alpha - \frac{|\mu_1 - \mu_0|}{\sigma/\sqrt{n}} \right].
\]

– We can repeat similar arguments for the left-tailed and two-tailed tests, obtaining the following formulae for power:

1. For a two-sided test:

\[
\text{Power} = P \left[ Z > z_\frac{\alpha}{2} - \frac{|\mu_0 - \mu_1|}{\sigma/\sqrt{n}} \right].
\]

2. For a one-sided test (left or right)

\[
\text{Power} = P \left[ Z > z_\alpha - \frac{|\mu_0 - \mu_1|}{\sigma/\sqrt{n}} \right].
\]

– Examples:
We wish to test the hypothesis that the mean weight for females who are 5'8" is 140 pounds. Assuming \( \sigma = 15 \) using a 5\% level of significance with \( n = 36 \), find the power of the test if \( \mu = 150 \) using a two-sided test.

**Solution:** Since this is two-tailed test,

\[
\text{Power} = P \left[ Z > z_{0.025} - \frac{|\mu_0 - \mu_1|}{\sigma/\sqrt{n}} \right].
\]

\( \mu_0 = 140, \mu_1 = 150, n = 36, \sigma = 15 \). \( z_{0.025} = 1.96 \), so that

\[
\text{Power} = P \left[ Z > 1.96 - \frac{140 - 150}{15/\sqrt{36}} \right] = P \left[ Z > -2.04 \right] = 1 - P \left[ Z < -2.04 \right] = 1 - 0.0207 = 0.9793.
\]

- **(Sample Size Determination)** Take again the right-tailed test. We have from the computation of power,

\[
1 - \beta = P \left[ Z > z_\alpha - \frac{|\mu_0 - \mu_1|}{\sigma/\sqrt{n}} \right].
\]

This defines,

\[
z_{1-\beta} = z_\alpha - \frac{|\mu_0 - \mu_1|}{\sigma/\sqrt{n}}.
\]

Since \( z_\beta = -z_{1-\beta} \),

\[
-z_\beta = z_\alpha - \frac{|\mu_0 - \mu_1|}{\sigma/\sqrt{n}} = z_\alpha - \frac{ES}{\sqrt{n}}
\]

so that,

\[
\sqrt{n} = \frac{z_\alpha + z_\beta}{ES} \iff n = \left( \frac{z_\alpha + z_\beta}{ES} \right)^2.
\]

- The same can be done for the left-tailed and two-tailed tests, so that we have the following formulae for sample-size determination,

1. For a two-tailed test,

\[
n = \left( \frac{z_{0.025} + z_\beta}{ES} \right)^2.
\]

2. For a one-sided test, it’s given by,

\[
n = \left( \frac{z_\alpha + z_\beta}{ES} \right)^2.
\]

- **Example:**
- We wish to run the following test:

\[ H_0 : \mu = 100 \]
\[ H_1 : \mu \neq 100 \]

at \( \alpha = 0.1 \). If \( \sigma = 20 \), how large of a sample would be required so that \( \beta = 0.04 \) if \( \mu = 102 \)?

**Solution:** This is a two-tailed test.

\[
ES = \frac{|\mu_0 - \mu_1|}{\sigma} = \frac{|100 - 102|}{20} = 0.1.
\]

Since \( \alpha = 0.1 \) and \( \beta = 0.04 \),

\[
n = \left( \frac{z_\alpha + z_\beta}{ES} \right)^2 = \left( \frac{z_{0.05} + z_{0.04}}{0.1} \right)^2 = \left( \frac{1.645 + 1.755}{0.1} \right)^2 = 34.
\]

Note: I looked up \( z_{0.05} \) and \( z_{0.04} \) in the z-table.