# Math C067 — Hypothesis Testing

Richard Beigel

April 21, 2006

**Example:** Checkout time at 7-11 is approximately normally distributed with mean 60 seconds and standard deviation 15 seconds. Management is considering a possible new checkout procedure. When they tested the new procedure on a random sample of 36 customers, they found the average waiting time was 55 seconds. How significant is this 5-second improvement? Is the new procedure responsible, or was the difference likely to have occured by chance?

To answer that question, we define **the null hypothesis** and several possible **alternative hypotheses**.

- Let $X_0$ be a random variable denoting the waiting time under the old procedure (known)
- Let $\overline{X}_0$ denote the sample mean of $X_0$ (known)
- Let $\mu_0 = E[X_0]$, the average waiting time under the old procedure (known)
- Let $X$ be a random variable denoting the waiting time under the new procedure
- Let $\mu = E[X]$, the average waiting time under the new procedure (unknown)
- Let $\bar{x}$, denote the average waiting time for our sample of the new procedure (known)

**Null hypothesis:** The new procedure makes no difference.

$H_0$**:** $\mu = \mu_0$

**Alternative hypotheses:** The new procedure decreases/increases/changes the waiting time.

$H_a$**:** $\mu < \mu_0$ (decrease)

$H_a$**:** $\mu > \mu_0$ (increase)

$H_a$**:** $\mu \neq \mu_0$ (increase or decrease)

**Very Important:** The alternative hypothesis must be formulated **before** the data is collected. Your alternative hypothesis can be based on existing data, but then **you must collect new data to test your hypothesis**.

We evaluate the new procedure by determining the P-value of the experimental results. The P-value is the probability that those results **or stronger results** would have resulted from the old system.

**The P-value is given by one of the formulas below, depending on the alternative hypothesis:**

$H_a$ **is** $\mu < \mu_0$**:**

$$P(\overline{X}_0 \leq \bar{x})$$

$H_a$ **is** $\mu > \mu_0$**:**

$$P(\overline{X}_0 \geq \bar{x})$$

$H_a$ **is** $\mu \neq \mu_0$**:**

$$\begin{cases} 2P(\overline{X}_0 \leq \bar{x}) & \text{if } \bar{x} \leq \mu_0 \\ 2P(\overline{X}_0 \geq \bar{x}) & \text{if } \bar{x} \geq \mu_0 \end{cases}$$

**Solved Example:** Checkout time at 7-11 is approximately normally distributed with mean 60 seconds and standard deviation 15 seconds. Management is considering a possible new checkout procedure.

(a) Formulate the null hypothesis and an alternative hypothesis.

When they tested the new procedure on a random sample of 36 customers, they found the average waiting time was 55 seconds.

(b) How significant is this 5-second improvement? Is the new procedure responsible, or was the difference likely to have occured by chance?

**Solution 1:**

(a) The null hypothesis is that the new average wait time $\mu = 60$. An alternative hypothesis is that $\mu < 60$.

(b) We need to calculate the P-value of the null hypothesis: $P(\overline{X}_0 \leq 55)$.

- The standard deviation of $\overline{X}_0$ is $15/\sqrt{36} = 15/6 = 2.5$
- 55 is $(60 - 55)/2.5 = 2$ standard deviations away from the mean.
- Looking up 2.00 in Table A-1 we find $P(55 < \overline{X}_0 \leq 60) \approx 0.47725$

$$\begin{aligned} P(\overline{X}_0 \leq 55) + P(55 < \overline{X}_0 \leq 60) &= P(\overline{X}_0 \leq 60) \\ P(\overline{X}_0 \leq 55) + P(55 < \overline{X}_0 \leq 60) &\approx 0.5 \\ P(\overline{X}_0 \leq 55) &\approx 0.5 - P(55 < \overline{X}_0 \leq 60) \\ P(\overline{X}_0 \leq 55) &\approx 0.5 - 0.47725 \\ P(\overline{X}_0 \leq 55) &\approx 0.02275 \end{aligned}$$

Thus the P-value of the null hypothesis is 0.02275. A P-value less than 0.05 is usually considered statistically significant, so we conclude the the reduced waiting times were unlikely to have been produced by mere chance and that the new checkout procedure is probably responsible for the reduced waiting times.

**Solution 2:**

- The null hypothesis is that the new average wait time $\mu = 60$. An alternative hypothesis is that $\mu > 60$.
- We need to calculate the P-value of the null hypothesis: $P(\overline{X}_0 \geq 55)$.
- The standard deviation of $\overline{X}_0$ is $15/\sqrt{36} = 15/6 = 2.5$
- 55 is $(60 - 55)/2.5 = 2$ standard deviations away from the mean.
- Looking up 2.00 in Table A-1 we find $P(55 \leq \overline{X}_0 \leq 60) \approx 0.47725$

$$\begin{aligned}
P(55 \leq \overline{X}_0) &= P(55 \leq \overline{X}_0 \leq 60) + P(60 < \overline{X}_0) \\
P(\overline{X}_0 \leq 55) &\approx 0.47725 + 0.5 \\
P(\overline{X}_0 \leq 55) &\approx 0.97725
\end{aligned}$$

Thus the P-value of the null hypothesis is 0.97725. Since the P-value is (much) greater than 0.05, we draw no conclusion. (The extremely large P-value indicates that the null hypothesis explains the experimental results better than the alternative hypothesis. However, it would be unwise to conclude that the null hypothesis is probably true. What actually happened is that the alternative hypothesis was really bad.)

Note: The P-values for the first two alternative hypotheses alway add up to 1.

**Solution 3:**

- The null hypothesis is that the new average wait time $\mu = 60$. An alternative hypothesis is that $\mu \neq 60$.
- We need to calculate the P-value of the null hypothesis: $2P(\overline{X}_0 \leq 55)$.
- The standard deviation of $\overline{X}_0$ is $15/\sqrt{36} = 15/6 = 2.5$
- 55 is $(60 - 55)/2.5 = 2$ standard deviations away from the mean.
- Looking up 2.00 in Table A-1 we find $P(55 < \overline{X}_0 \leq 60) \approx 0.047725$

$$
\begin{aligned}
P(\overline{X}_0 \leq 55) + P(55 < \overline{X}_0 \leq 60) &= P(\overline{X}_0 \leq 60) \\
P(\overline{X}_0 \leq 55) + P(55 < \overline{X}_0 \leq 60) &\approx 0.5 \\
P(\overline{X}_0 \leq 55) &\approx 0.5 - P(55 < \overline{X}_0 \leq 60) \\
P(\overline{X}_0 \leq 55) &\approx 0.5 - 0.47725 \\
P(\overline{X}_0 \leq 55) &\approx 0.02275 \\
2P(\overline{X}_0 \leq 55) &\approx 0.0455
\end{aligned}
$$

Thus the P-value of the null hypothesis is 0.0455. A P-value less than 0.05 is usually considered statistically significant, so we conclude the the reduced waiting times were unlikely to have been produced by mere chance and that the new checkout procedure is probably responsible for the reduced waiting times.

Note: The P-value for the third alternative hypothesis is always equal to twice the P-value for one of the first two alternative hypotheses (the first, in this example).

**Exercise 1: Popcorn, no control group** A medical student theorizes that popcorn causes cancer in lab rats. Assume that the overall cancer rate among lab rats is known to be 10%. A test group of 10 rats has popcorn mixed into their ordinary diet, and 2 rats in the test group develop cancer.

(a) Formulate the null hypothesis and the med student's alternative hypothesis.

(b) What is the P-value for the med student's experiment?

(c) Does the med student's experiment significantly support the theory that popcorn causes cancer in lab rats?

**Solution:**

- Let $p_0$ denote the proportion of rats developing cancer in the general rat population on the standard diet. As given, $p_0 = 0.1$.
- Let $p$ denote the (unknown) probability that a rat would develop cancer on the popcorn-added diet
- The null hypothesis is $p = p_0$, i.e., $p = 0.1$
- The med student's alternative hypothesis is $p > 0.1$
- Let $n$ denote the size of the test group (sample size). As given, $n = 10$.
- Let $\hat{P}$ denote the sample proportion of cancer in $n$ randomly chosen rats on the **standard diet**
- Then $E[\hat{P}] = p = 0.1$
- and $\mathrm{Var}[\hat{P}] = p(1-p)/n = 0.1 \cdot 0.9/10 = 0.009$ (I got a different value in class. What was my mistake?)
- so $\sigma = \sqrt{0.009} \approx 0.90513167$.
- Let $\hat{p}$ denote the measured proportion of rats who developed cancer in our test group.
- Then $\hat{p} = 2/10 = 0.2$
- $\hat{p} - p = 0.2 - 0.1 = 0.1$, which consists of $0.1/0.90513167 \approx 1.05409255$ standard deviations
- The number of degrees of freedom is $k = n - 1 = 10 - 1 = 9$
- We look for 1.05 in row 9 of Table A-2:
- We find 0.883 in column 0.3 and we find 1.38 in column 0.4
- 1.05 belongs between columns 0.3 and 0.4
- So the P-value is between $0.5 - 0.4 = 0.1$ and $0.5 - 0.3 = 0.2$
- Since the P-value is greater than 0.05, we do not accept or reject the null hypothesis.

**Exercise 2: Hot dogs, no control group** A second medical student theorizes that hot dogs cause cancer in lab rats. Assume that the overall cancer rate among lab rats is known to be 10%. A test group of 100 rats has hot dogs mixed into their ordinary diet, and 20 rats in the test group develop cancer.

(a) Formulate the null hypothesis and the second med student's alternative hypothesis.

(b) What is the P-value for the second med student's experiment?

(c) Does the med student's experiment significantly support the theory that hot dogs cause cancer in lab rats?

Use Table A-1 for this problem because the sample size is large.

**Example 1′: Popcorn, control group** A medical student theorizes that popcorn causes cancer in lab rats. A control group of 200 rats is fed their ordinary diet, and a test group of 10 rats has popcorn mixed into their ordinary diet. 20 rats in the control group develop cancer and 2 rats in the test group develop cancer.

(a) Formulate the null hypothesis and the med student's alternative hypothesis.

(b) What is the P-value for the med student's experiment?

(c) Does the med student's experiment significantly support the theory that popcorn causes cancer in lab rats?

**Solution:**

- Let $p_1$ denote the proportion of rats in the population who would develop cancer on their ordinary diet
- Let $p_2$ denote the proportion of rats in the population who would develop cancer if popcorn was added to their diet
- The null hypothesis is $p_2 = p_1$, i.e., $p_2 - p_1 = 0$
- The med student's alternative hypothesis is $p_2 > p_1$, i.e., $p_2 - p_1 > 0$
- Let $n_1$ denote the size of the control group, and $n_2$ the size of the test group
- Let $\hat{P}_1$ denote the proportion of cancer in a random sample of $n_1$ rats on an ordinary diet
- Let $\hat{P}_2$ denote the proportion of cancer in a random sample of $n_2$ rats on the popcorn-added diet
- Let $\hat{p}_1$ and $\hat{p}_2$ denote the measured values of $\hat{P}_1$ and $\hat{P}_2$
- Let $X = \hat{P}_2 - \hat{P}_1$
- Then $\mu_X = E[X] = E[\hat{P}_2] - E[\hat{P}_1]$ so we estimate

$$\mu_X \approx \hat{p}_2 - \hat{p}_1 = 0.2 - 0.1 = 0.1$$

- and $\mathrm{Var}[X] = \mathrm{Var}[\hat{P}_2] + \mathrm{Var}[\hat{P}_1]$ so we estimate

$$\begin{aligned}
\mathrm{Var}[X] &\approx \frac{\hat{p}_2(1-\hat{p}_2)}{n_2} + \frac{\hat{p}_1(1-\hat{p}_1)}{n_1} \\
\mathrm{Var}[X] &\approx \frac{0.2 \cdot 0.8}{10} + \frac{0.1 \cdot 0.9}{200} \\
\mathrm{Var}[X] &\approx 0.016 + 0.00045 \\
\mathrm{Var}[X] &\approx 0.01645 \\
\sigma_X &\approx 0.128
\end{aligned}$$

- Thus we estimate that $p_2 - p_1$ consists of $0.1/0.128 \approx 0.78$ standard deviations.
- The number of degrees of freedom is given by the following formula

$$\begin{aligned} k &= \frac{\left(\frac{\hat{p}_2(1-\hat{p}_2)}{n_2} + \frac{\hat{p}_1(1-\hat{p}_1)}{n_1}\right)^2}{\frac{1}{n_2-1}\left(\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}\right)^2 + \frac{1}{n_1-1}\left(\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}\right)^2} \\ &= \frac{0.01645^2}{\frac{1}{9}0.016^2 + \frac{1}{199}0.00045^2} \\ &= 9.51 \end{aligned}$$

  Rounding *down*, we take $k = 9$.

  **Note:** *When the control group is much larger than the test group, the number of degrees of freedom will be close to* $n_2 - 1$*. You can use this rule of thumb to check your calculation. If you are short on time, you can skip the calculation and use this rule for partial credit.*

- We look for 0.78 in row 9 of Table A-2
- 0.78 is between 0.703 **in column 0.25** and 0.883 **in column 0.3**
- Therefore

$$\begin{array}{rcccl} 0.25 & \le & P(0 \le X \le 0.1) & \le & 0.3 \\ 0.5 - 0.25 & \ge & 0.5 - P(0 \le X < 0.1) & \ge & 0.5 - 0.3 \\ 0.25 & \ge & P(0 \le X) - P(0 \le X < 0.1) & \ge & 0.2 \\ 0.2 & \le & P(0 \le X) - P(0 \le X < 0.1) & \le & 0.5 \\ 0.2 & \le & P(0.1 \le X) & \le & 0.25 \\ 0.2 & \le & P(X \ge 0.1) & \le & 0.25 \end{array}$$

  Therefore the P-value is between 0.2 and 0.25.

- Since the P-value is at least 0.2, which is greater than 0.05, the test results are not significant and we do not conclude anything about whether popcorn causes cancer in lab rats.

**Example 2′: Hot dogs, control group** A medical student theorizes that hot dogs cause cancer in lab rats. A control group of 200 rats is fed their ordinary diet, and a test group of 100 rats has hot dogs mixed into their ordinary diet. 20 rats in the control group develop cancer and 20 rats in the test group develop cancer.

(a) Formulate the null hypothesis and the second med student's alternative hypothesis.

(b) What is the P-value for the second med student's experiment?

(c) Does the med student's experiment significantly support the theory that hot dogs cause cancer in lab rats?

**Solution:**

- Let $p_1$ denote the proportion of rats in the population who would develop cancer on their ordinary diet
- Let $p_2$ denote the proportion of rats in the population who would develop cancer if popcorn was added to their diet
- The null hypothesis is $p_2 = p_1$, i.e., $p_2 - p_1 = 0$
- The med student's alternative hypothesis is $p_2 > p_1$, i.e., $p_2 - p_1 > 0$
- Let $n_1$ denote the size of the control group, and $n_2$ the size of the test group
- Let $\hat{P}_1$ denote the proportion of cancer in a random sample of $n_1$ rats on an ordinary diet
- Let $\hat{P}_2$ denote the proportion of cancer in a random sample of $n_2$ rats on the popcorn-added diet
- Let $\hat{p}_1$ and $\hat{p}_2$ denote the measured values of $\hat{P}_1$ and $\hat{P}_2$
- Let $X = \hat{P}_2 - \hat{P}_1$
- Then $\mu_X = E[X] = E[\hat{P}_2] - E[\hat{P}_1]$ so we estimate

$$\mu_X \approx \hat{p}_2 - \hat{p}_1 = 0.2 - 0.1 = 0.1$$

- and $\text{Var}[X] = \text{Var}[\hat{P}_2] + \text{Var}[\hat{P}_1]$ so we estimate

$$\begin{aligned}
\text{Var}[X] &\approx \frac{\hat{p}_2(1-\hat{p}_2)}{n_2} + \frac{\hat{p}_1(1-\hat{p}_1)}{n_1} \\
\text{Var}[X] &\approx \frac{0.2 \cdot 0.8}{100} + \frac{0.1 \cdot 0.9}{200} \\
\text{Var}[X] &\approx 0.0016 + 0.00045 \\
\text{Var}[X] &\approx 0.00205 \\
\sigma_X &\approx 0.0452769257
\end{aligned}$$

- Thus we estimate that $p_2 - p_1$ consists of $0.1/0.0452769257 \approx 2.21$ standard deviations.
- Because the control group and the test group are both large (in general you should check that the the number of positives is at least 5 and the number of negatives is at least 5 both in the control group and in the test group), we can use Table A-1 instead of calculating degrees of freedom.
- To look up 2.21 in Table A-1 we look in row 2.2 column 1 to find $P(0 \leq X < 0.1) \approx 0.4864$.
- We have

$$\begin{aligned}
P(0 \leq X < 0.1) + P(0.1 \leq X) &= P(0 \leq X) \\
P(0 \leq X < 0.1) + P(0.1 \leq X) &\approx 0.5 \\
0.4864 + P(0.1 \leq X) &\approx 0.5 \\
P(0.1 \leq X) &\approx 0.5 - 0.4864 \\
P(X \geq 0.1) &\approx 0.0136
\end{aligned}$$

Since the P-value is less than 0.05, the test results are significant, and we reject the null hypothesis.