7. Hypothesis Testing

Dave Goldsman

H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

1/15/20

Georgia Tech

Outline

Lesson 7.1 — Introduction to Hypothesis Testing

Goal: In this module, we'll study a population by collecting data and making sound, statistically valid conclusions about that population based on data that we collect.

General Approach

1. State a hypothesis.

2. Select a test statistic (to test whether or not the hypothesis is true).

3. Evaluate (calculate) the test statistic based on observations that we take.

4. Interpret results — does the test statistic suggest that you reject or fail to reject your hypothesis?

Details follow. . . .

1. **Hypotheses** are simply statements or claims about parameter values.

You perform a **hypothesis test** to prove or disprove the claim.

Set up a **null hypothesis (H_0)** and an **alternative hypothesis (H_1)** to cover the entire parameter space. The null hypothesis sort of represents the "status quo." It's not necessarily true, but we will grudgingly stick with it until proven otherwise.

Example: We currently believe that the mean weight of a filled package of chicken is μ_0 ounces. (We specify μ_0.) But we have our suspicions.

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

This is a **two-sided test**. We will reject the belief of the null hypothesis H_0 if $\hat{\mu}$ (an estimator of μ) is "too high" or "too small."

Example: We hope that a new brand of tires will last for a mean of more than μ_0 miles. (We specify μ_0.) But we really need evidence before we can state that claim with reasonable certainty. Else, we'll stay with the old brand.

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

This is a **one-sided test**. We'll reject H_0 only if $\hat{\mu}$ is "too large."

Example: We test to see if emissions from a certain type of car are less than a mean of μ_0 ppm. But we need evidence.

$$H_0 : \mu \geq \mu_0$$
$$H_1 : \mu < \mu_0$$

This is a **one-sided test**. We'll reject the null hypothesis if $\hat{\mu}$ is "too small," and only then will make the claim that the emissions are low.

Idea: H_0 is the old, conservative "status quo." H_1 is the new, radical hypothesis. Although you may not be tooooo sure about the truth of H_0, you won't reject it in favor of H_1 unless you see substantial evidence in support of H_1.

Think of H_0 as "Innocent until proven guilty."

If you get substantial evidence supporting H_1, you'll decide to reject H_0. Otherwise, you "fail to reject" H_0. (This roughly means that you grudgingly accept H_0.)

2. Select a **test statistic** (a random variable that we'll use to test if H_0 is true).

For instance, we could compare an estimator $\hat{\mu}$ with μ_0. The comparison is accomplished using a known sampling distribution (aka "test statistic"), e.g.,

$$z_{\text{obs}} = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \quad \text{(if } \sigma^2 \text{ is known) or}$$

$$t_{\text{obs}} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \quad \text{(if } \sigma^2 \text{ is unknown).}$$

Lots more details later.

3. Evaluate the test statistic. Here's the logic of hypothesis testing:

    (a) Collect sample data.

    (b) Calculate the value of the test statistic based on the data.

    (c) Assume H_0 (the "status quo") is true.

    (d) Determine the probability of the sample result, assuming H_0 is true.

    (e) Decide from (d) if H_0 is plausible:

    – If the probability from (d) is low, reject H_0 and select H_1.
    – If the probability from (d) is high, fail to reject H_0.

Example: Time to metabolize a drug. The current drug takes $\mu_0 = 15$ min. Is the new drug better?

Claim: Expected time for new drug is < 15 min.

$$H_0 : \mu \geq 15$$
$$H_1 : \mu < 15$$

Data: $n = 20$, $\bar{X} = 14.88$, $S = 0.333$.

The test statistic is

$$t_{\text{obs}} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = -1.61.$$

Now, if H_0 is actually the true state of things, then $\mu = \mu_0$, and from our discussion on CI's, we have

$$t_{\text{obs}} \;=\; \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \;\sim\; t(n-1) \;\sim\; t(19).$$

What would cause us to reject H_0?

If $\bar{X} \ll \mu_0 (= 15)$, this would indicate that H_0 is probably wrong.

Equivalently, I'd reject H_0 if t_{obs} is "significantly" $\ll 0$.

4. Interpret the Test Statistic.

So if H_0 is true, is it reasonable (or, at least, not outrageous) to have gotten $t_{\text{obs}} = -1.61$?

If yes, then we we'll *fail to reject* ("grudgingly accept") H_0.

If no, then we'll *reject* H_0 in favor of H_1.

Let's see. . . . From the t table, we have

$$t_{0.95,19} = -t_{0.05,19} = -1.729 \quad \text{and} \quad t_{0.90,19} = -t_{0.10,19} = -1.328,$$

i.e.,

$$\begin{aligned} P(t(19) < -1.729) &= 0.05 \ \text{ and} \\ P(t(19) < -1.328) &= 0.10. \end{aligned}$$

This means that

$$0.05 \ < \ p \ \equiv \ P(t(19) < \underbrace{-1.61}_{t_{\text{obs}}}) \ < \ 0.10.$$

In English: If H_0 were true, there's a $100p\%$ chance that we'd see a value of t_{obs} that's ≤ -1.61. That's not a real high probability, but it's not toooo small.

Formally, we'd **reject** H_0 at "level" 0.10, since $t_{\text{obs}} = -1.61 < -t_{0.10,19} = -1.328$.

But, we'd **fail to reject** H_0 at level 0.05, since $t_{\text{obs}} = -1.61 > -t_{0.05,19} = -1.729$.

More on this pretty soon! $\square$

Outline

Lesson 7.2 — The Errors of Our Ways

Where Can We Go Wrong? Four things can happen:

- If H_0 is actually true and we conclude that it's true — good.
- If H_0 is actually false and we conclude that it's false — good.
- If H_0 is actually true and we conclude that it's false — bad. This is called **Type I error**.
- If H_0 is actually false and we conclude that it's true — bad. This is called **Type II error**.

State of nature	Decision: Accept H_0	Decision: Reject H_0
H_0 true	Correct! ☺	Type I error ☹
H_0 false	Type II error ☹	Correct! ☺

Example: We incorrectly conclude that a new, inferior drug is better than the drug currently on the market — Type I error.

Example: We incorrectly conclude that a new, superior drug is worse than the drug currently on the market — Type II error.

We want to keep:

$$\begin{aligned} P(\text{Type I error}) &= P(\text{Reject } H_0 \mid H_0 \text{ true}) \ \le \ \alpha. \\ P(\text{Type II error}) &= P(\text{Fail to Rej } H_0 \mid H_0 \text{ false}) \ \le \ \beta. \end{aligned}$$

We choose α and β. Of course, we need to have $\alpha + \beta < 1$.

Usually, Type I error is considered to be "worse" than Type II.

Definition: The probability of Type I error, α, is called the **size** or **level of significance** of the test.

Definition: The **power** of a hypothesis test is

$$1 - \beta \ = \ P(\text{Reject } H_0 \mid H_0 \text{ false}).$$

It's good to have high power.

Lesson 7.3 — Normal Mean Test with Known Variance

We'll discuss hypothesis tests involving the mean(s) of normal distribution(s) under various scenarios, all of which involve *known variance(s)*.

Suppose that $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} \text{Nor}(\mu, \sigma^2)$, where σ^2 is somehow *known* (which is not very realistic).

Two-sided test (also known as a **simple test**):

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_1 : \mu \neq \mu_0.$$

We'll use $\bar{X}$ to estimate μ. If $\bar{X}$ is "significantly different" than μ_0, then we'll reject H_0. But how much is "significantly different"?

To determine what "significantly different" means, first define

$$Z_0 \;\equiv\; \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}.$$

If H_0 is true, then $\mathrm{E}[\bar{X}] = \mu_0$ and $\mathrm{Var}(\bar{X}) = \sigma^2/n$; and so $Z_0 \sim \mathrm{Nor}(0,1)$.

Then we have

$$P(-z_{\alpha/2} \le Z_0 \le z_{\alpha/2}) \;=\; 1-\alpha.$$

A value of Z_0 outside the interval $[-z_{\alpha/2}, z_{\alpha/2}]$ is highly unlikely if H_0 is true.

Therefore,

$$\text{Reject } H_0 \quad \text{iff} \quad |Z_0| > z_{\alpha/2}.$$

This assures us that

$$\begin{aligned} P(\text{Type I error}) &= P(\text{Reject } H_0 \mid H_0 \text{ true}) \\ &= P\big(|Z_0| > z_{\alpha/2} \mid Z_0 \sim \text{Nor}(0,1)\big) \\ &= \alpha. \end{aligned}$$

If $|Z_0| > z_{\alpha/2}$, then we're in the **rejection region**.
(This is also called the **critical region**.)

If $|Z_0| \le z_{\alpha/2}$, then we're in the **acceptance region**.

One-sided test:

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

Again let

$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}.$$

A value of Z_0 outside the interval $(-\infty, z_\alpha]$ is highly unlikely if H_0 is true. Therefore,

$$\text{Reject } H_0 \quad \text{iff} \quad Z_0 > z_\alpha.$$

If $Z_0 > z_\alpha$, this suggests $\mu > \mu_0$.

Similarly, the other one-sided test:

$$H_0 : \mu \geq \mu_0$$
$$H_1 : \mu < \mu_0$$

A value of Z_0 outside the interval $[-z_\alpha, \infty)$ is highly unlikely if H_0 is true. Therefore,

$$\text{Reject } H_0 \quad \text{iff} \quad Z_0 < -z_\alpha.$$

If $Z_0 < -z_\alpha$, this suggests $\mu < \mu_0$.

Example: We examine the weights of 25 nine-year-old kids. Suppose we somehow know that the weights are normally distributed with $\sigma = 4$. The sample mean of the 25 weights is 62.

Test the hypothesis that the mean weight is 60.
Keep the probability of Type I error $= 0.05$.

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_1 : \mu \neq \mu_0.$$

Here we have

$$Z_0 \;=\; \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \;=\; \frac{62 - 60}{4/\sqrt{25}} \;=\; 2.5.$$

Since $|Z_0| = 2.5 > z_{\alpha/2} = z_{0.025} = 1.96$, we *reject* H_0.

Notice that a lower α results in a higher $z_{\alpha/2}$. Then it's "harder" to reject H_0. For instance, if $\alpha = 0.01$, then $z_{0.005} = 2.58$, and we would *fail to reject* H_0 in this example. $\square$

Definition: The ***p*-value** of a test is the smallest level of significance α that would lead to rejection of H_0.

Remark: Researchers often report the p-values of any tests that they conduct.

For the two-sided normal mean test with known variance, we reject H_0 iff

$$|Z_0| \;>\; z_{\alpha/2} \;=\; \Phi^{-1}(1-\alpha/2)$$

$$\text{iff} \quad \Phi(|Z_0|) \;>\; 1-\alpha/2$$

$$\text{iff} \quad \alpha \;>\; 2(1-\Phi(|Z_0|)).$$

Thus, for the two-sided test

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_1 : \mu \neq \mu_0,$$

the p-value is $p = 2(1-\Phi(|Z_0|))$. $\square$

Similarly, for the one-sided test

$$H_0 : \mu \leq \mu_0 \quad \text{vs.} \quad H_1 : \mu > \mu_0,$$

we have $p = 1 - \Phi(Z_0)$.

And for the other one-sided test

$$H_0 : \mu \geq \mu_0 \quad \text{vs.} \quad H_1 : \mu < \mu_0,$$

we have $p = \Phi(Z_0)$.

Example: For the previous example,

$$p \;=\; 2(1 - \Phi(|Z_0|)) \;=\; 2(1 - \Phi(2.5)) \;=\; 0.0124. \quad \square$$

Lesson 7.4 — Normal Mean Test with Known Variance: Design

Goal: Design a two-sided test with the following constraints on Type I and II errors:

$$P(\text{Type I error}) \ \leq \ \alpha \quad \text{and} \quad P\big(\text{Type II error} \,|\, \mu = \mu_1 > \mu_0\big) \ \leq \ \beta.$$

By "design," we mean how many observations n do we need for the two-sided test to satisfy a Type I error bound of α and a Type II error bound of β?

Remark: The bound β is for the *special case* that the true mean μ happens to equal a *user-specified* value $\mu = \mu_1 > \mu_0$. In other words, we're trying to "protect" ourselves against the possibility that μ actually happens to equal μ_1.

If we change the "protected" μ_1, we'll need to change n. Generally, the closer μ_1 is to μ_0, the more work we need to do (i.e., higher n) — because it's harder to distinguish between two close cases.

Theorem: Suppose the difference between the actual and hypothesized means is

$$\delta \equiv \mu - \mu_0 = \mu_1 - \mu_0.$$

(Without loss of generality, we'll assume $\mu_1 > \mu_0$.) Then the α and β design requirements can be achieved by taking a sample of size

$$n \doteq \sigma^2 (z_{\alpha/2} + z_\beta)^2 / \delta^2.$$

Remark: In the proof that follows, we'll get an expression for β that involves the standard normal cdf evaluated at a mess that contains n. We'll then do an inversion to obtain the desired approximation for n.

Proof: Let's first look at the β value,

$$
\begin{aligned}
\beta &= P(\text{Type II error} \,|\, \mu = \mu_1 > \mu_0) \\
&= P(\text{Fail to Reject } H_0 \,|\, H_0 \text{ false } (\mu = \mu_1 > \mu_0)) \\
&= P(|Z_0| \le z_{\alpha/2} \,|\, \mu = \mu_1) \\
&= P(-z_{\alpha/2} \le Z_0 \le z_{\alpha/2} \,|\, \mu = \mu_1) \\
&= P\left(-z_{\alpha/2} \le \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \le z_{\alpha/2} \,\middle|\, \mu = \mu_1\right) \\
&= P\left(-z_{\alpha/2} \le \frac{\bar{X} - \mu_1}{\sigma/\sqrt{n}} + \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} \le z_{\alpha/2} \,\middle|\, \mu = \mu_1\right).
\end{aligned}
$$

Notice that

$$Z \;\equiv\; \frac{\bar{X} - \mu_1}{\sigma/\sqrt{n}} \;\sim\; \text{Nor}(0,1).$$

This gives

$$\begin{aligned}
\beta &= P\left(-z_{\alpha/2} \;\le\; Z + \frac{\sqrt{n}\delta}{\sigma} \;\le\; z_{\alpha/2}\right) \\
&= P\left(-z_{\alpha/2} - \frac{\sqrt{n}\delta}{\sigma} \;\le\; Z \;\le\; z_{\alpha/2} - \frac{\sqrt{n}\delta}{\sigma}\right) \\
&= \Phi\left(z_{\alpha/2} - \frac{\sqrt{n}\delta}{\sigma}\right) - \Phi\left(-z_{\alpha/2} - \frac{\sqrt{n}\delta}{\sigma}\right).
\end{aligned}$$

Now, note that $-z_{\alpha/2} \ll 0$ and $-\sqrt{n}\delta/\sigma < 0$ (since $\delta > 0$).

These two facts imply that the second $\Phi(\cdot)$ is pretty much zero, so…

We only need to use the first term in the previous expression for β:

$$\beta \doteq \Phi\left(z_{\alpha/2} - \frac{\sqrt{n}\delta}{\sigma}\right)$$

iff

$$\Phi^{-1}(\beta) = -z_\beta \doteq z_{\alpha/2} - \frac{\sqrt{n}\delta}{\sigma}$$

iff

$$\frac{\sqrt{n}\delta}{\sigma} \doteq z_{\alpha/2} + z_\beta$$

iff

$$n \doteq \sigma^2(z_{\alpha/2} + z_\beta)^2/\delta^2. \quad \text{Done! Whew!}$$

Recap: If you want to test $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$, and

(1) You know σ^2,

(2) You want $P(\text{Type I error}) = \alpha$, and

(3) You want $P(\text{Type II error}) = \beta$ when $\mu = \mu_1 (\neq \mu_0)$,

then you have to take $n \doteq \sigma^2 (z_{\alpha/2} + z_\beta)^2 / \delta^2$ observations.

Similarly, if you're doing a *one-sided* test, it turns out that you need to take $n \doteq \sigma^2 (z_\alpha + z_\beta)^2 / \delta^2$ observations.

Example: Weights of 9-year-old kids are normal with $\sigma = 4$. How many observations should we take if we wish to test $H_0 : \mu = 60$ vs. $H_1 : \mu \neq 60$, and we want $\alpha = 0.05$ and $\beta = 0.05$, if μ happens to actually equal $\mu_1 = 62$?

$$n \doteq \frac{\sigma^2}{\delta^2}(z_{\alpha/2} + z_\beta)^2 = \frac{16}{4}(1.96 + 1.645)^2 = 51.98.$$

In other words, we need about 52 observations. $\square$

Lesson 7.5 — Two-Sample Normal Means Test with Known Variances

Suppose we have the following set-up:

$$X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Nor}(\mu_x, \sigma_x^2) \quad \text{and}$$

$$Y_1, Y_2, \ldots, Y_m \overset{\text{iid}}{\sim} \text{Nor}(\mu_y, \sigma_y^2),$$

where the samples are independent of each other,
and σ_x^2 and σ_y^2 are somehow *known*.

Which of the two populations has the larger mean?

Here's the two-sided test to see if the means are different.

$$H_0 : \mu_x = \mu_y$$
$$H_1 : \mu_x \neq \mu_y$$

Define the test statistic

$$Z_0 \;=\; \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}.$$

If H_0 is true (i.e., the means are equal), then

$$Z_0 \;=\; \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \;\sim\; \text{Nor}(0,1).$$

Thus, as before,

$$\text{Reject } H_0 \quad \text{iff} \quad |Z_0| > z_{\alpha/2}.$$

Using more of the same reasoning as before, we get the following one-sided tests:

$$H_0 : \mu_x \le \mu_y \quad \text{vs.} \quad H_1 : \mu_x > \mu_y.$$

$$\text{Reject } H_0 \quad \text{iff} \quad Z_0 \;>\; z_\alpha.$$

$$H_0 : \mu_x \ge \mu_y \quad \text{vs.} \quad H_1 : \mu_x < \mu_y.$$

$$\text{Reject } H_0 \quad \text{iff} \quad Z_0 \;<\; -z_\alpha.$$

It's so easy!! ☺

Example: Suppose we want to test $H_0 : \mu_x = \mu_y$ vs. $H_1 : \mu_x \neq \mu_y$, and we have the following data:

$$\begin{aligned} n &= 10, \quad \bar{X} = 824.9, \quad \sigma_x^2 = 40 \\ m &= 10, \quad \bar{Y} = 818.6, \quad \sigma_y^2 = 50, \end{aligned}$$

where the variances are somehow known.

Then

$$Z_0 = \frac{824.9 - 818.6}{\sqrt{\frac{40}{10} + \frac{50}{10}}} = 2.10.$$

If $\alpha = 0.05$, then $|Z_0| = 2.10 > z_{\alpha/2} = 1.96$, and so we *reject* H_0. $\square$

Lesson 7.6 — Normal Mean Test with Unknown Variance

The next few lessons deal with the *unknown* variance case.

- Test the mean of a single normal distribution (here).
- Compare the means of two normal distributions when both variances are unknown (the following three lessons deal with different subcases).

Anyhow, it's time for t again!

Suppose $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} \text{Nor}(\mu, \sigma^2)$, where σ^2 is *unknown*.

Two-sided (aka simple) test for one normal population:

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_1 : \mu \neq \mu_0.$$

We'll use $\bar{X}$ to estimate μ. If $\bar{X}$ is "significantly different" than μ_0, then we'll reject H_0. For this purpose, we'll also need to estimate σ^2.

Define the test statistic

$$T_0 \equiv \frac{\bar{X} - \mu_0}{S/\sqrt{n}},$$

where S^2 is our old friend, the sample variance,

$$S^2 \equiv \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \sim \frac{\sigma^2\chi^2(n-1)}{n-1}.$$

If H_0 is true, then

$$T_0 = \frac{\frac{\bar{X}-\mu_0}{\sqrt{\sigma^2/n}}}{\sqrt{S^2/\sigma^2}} \sim \frac{\mathrm{Nor}(0,1)}{\sqrt{\frac{\chi^2(n-1)}{n-1}}} \sim t(n-1).$$

So the **two-sided test** is:

$$\text{Reject } H_0 \quad \text{iff} \quad |T_0| > t_{\alpha/2,n-1}.$$

Using the same reasoning as in previous lessons, the **one-sided tests** are:

$$H_0 : \mu \le \mu_0 \quad \text{vs.} \quad H_1 : \mu > \mu_0.$$

$$\text{Reject } H_0 \quad \text{iff} \quad T_0 \; > \; t_{\alpha,n-1}.$$

$$H_0 : \mu \ge \mu_0 \quad \text{vs.} \quad H_1 : \mu < \mu_0.$$

$$\text{Reject } H_0 \quad \text{iff} \quad T_0 \; < \; -t_{\alpha,n-1}.$$

Recall: The **p-value** of a test is the smallest level of significance α that would lead to rejection of H_0.

For this two-sided normal mean test with unknown variance, we reject H_0 iff

$$|T_0| \; > \; t_{\alpha/2,n-1} \; = \; F_{n-1}^{-1}(1-\alpha/2),$$

where $F_{n-1}(t)$ is the cdf of the $t(n-1)$ distribution (and $F_{n-1}^{-1}(\cdot)$ is the inverse). This relationship holds iff

$$F_{n-1}(|T_0|) \; > \; 1-\alpha/2 \quad \text{iff} \quad \alpha \; > \; 2\big(1-F_{n-1}(|T_0|)\big).$$

Thus, for the two-sided test for the case of unknown variance,

$$H_0 : \mu = \mu_0 \quad \text{vs.} \quad H_1 : \mu \neq \mu_0,$$

the p-value is $p = 2\big(1-F_{n-1}(|T_0|)\big)$. $\square$

Example: Suppose we want to test at level 0.05 whether or not the mean of some process is 150.

Data: $n = 15$, $\bar{X} = 152.18$, and $S^2 = 16.63$.

Then
$$T_0 \equiv \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = 2.07.$$

Let's do a two-sided test at level $\alpha = 0.05$. Since $|T_0| < t_{\alpha/2,n-1} = t_{0.025,14} = 2.145$, we (barely) *fail to reject* H_0.

Alternatively, note that the p-value is $2\big(1 - F_{14}(2.07)\big) = 0.0574$, where the Excel function `T.DIST` can be used to evaluate the cdf $F_{14}(2.07)$. Since $p = 0.0574 > 0.05 = \alpha$, we didn't quite reject. $\Box$

Lesson 7.7 — Two-Sample Normal Means Tests with Unknown Variances

Suppose we have the following set-up:

$$\begin{aligned} X_1, X_2, \ldots, X_n &\overset{\text{iid}}{\sim} \text{Nor}(\mu_x, \sigma_x^2) \\ Y_1, Y_2, \ldots, Y_m &\overset{\text{iid}}{\sim} \text{Nor}(\mu_y, \sigma_y^2), \end{aligned}$$

where the samples are independent of each other, and σ_x^2 and σ_y^2 are *unknown*.

Which population has the larger mean?

We'll look at three cases:

Pooled t-test: $\sigma_x^2 = \sigma_y^2 = \sigma^2$ (next, this lesson).

Approximate t-test: $\sigma_x^2 \neq \sigma_y^2$ (later, this lesson).

Paired t-test: (X_i, Y_i) observations paired (next lesson).

Pooled t-Test

Suppose that $\sigma_x^2 = \sigma_y^2 = \sigma^2$ (unknown).

Consider the two-sided test to see if the means are different,

$$H_0 : \mu_x = \mu_y \quad \text{vs.} \quad H_1 : \mu_x \neq \mu_y.$$

Sample means and variances from the two populations,

$$\bar{X} \equiv \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad \bar{Y} \equiv \frac{1}{m}\sum_{i=1}^{m} Y_i$$

$$S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \quad \text{and} \quad S_y^2 = \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \bar{Y})^2.$$

As in the previous module, define the **pooled variance estimator** by

$$S_p^2 \equiv \frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}.$$

If H_0 is true, it can be shown that

$$S_p^2 \sim \frac{\sigma^2\chi^2(n+m-2)}{n+m-2},$$

and then the test statistic

$$T_0 \equiv \frac{\bar{X}-\bar{Y}}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim t(n+m-2).$$

Thus,

$$\text{Reject } H_0 \quad \text{iff} \quad |T_0| > t_{\alpha/2,n+m-2}.$$

One-Sided Tests:

$$H_0 : \mu_x \le \mu_y \quad \text{vs.} \quad H_1 : \mu_x > \mu_y.$$

$$\text{Reject } H_0 \quad \text{iff} \quad T_0 \ > \ t_{\alpha,n+m-2}.$$

$$H_0 : \mu_x \ge \mu_y \quad \text{vs.} \quad H_1 : \mu_x < \mu_y.$$

$$\text{Reject } H_0 \quad \text{iff} \quad T_0 \ < \ -t_{\alpha,n+m-2}.$$

Example: Catalyst X is currently used by a certain chemical process. If catalyst Y gives higher mean yield, we'll use it instead.

Thus, we want to test $H_0 : \mu_x \ge \mu_y$ vs. $H_1 : \mu_x < \mu_y$.

Suppose we have the following data:

$$\begin{aligned} n &= 8, \quad \bar{X} = 91.73, \quad S_x^2 = 3.89 \\ m &= 8, \quad \bar{Y} = 93.75, \quad S_y^2 = 4.02. \end{aligned}$$

S_x^2 is pretty close to S_y^2, so we'll assume $\sigma_x^2 \doteq \sigma_y^2$.

This justifies the use of the pooled variance estimator

$$S_p^2 \;=\; \frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2} \;=\; 3.955,$$

so that

$$T_0 \;=\; \frac{\bar{X} - \bar{Y}}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \;=\; -2.03.$$

Let's test at level $\alpha = 0.05$. Then

$$t_{\alpha,n+m-2} \;=\; t_{0.05,14} \;=\; 1.761.$$

Since $T_0 < -t_{\alpha,n+m-2}$, we *reject* H_0.

Thus, we should probably use catalyst Y. $\square$

Approximate t-Test

Suppose that $\sigma_x^2 \neq \sigma_y^2$ (both unknown). As with our work with CIs, define

$$T_0^\star \equiv \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} \approx t(\nu) \quad \text{(if } H_0 \text{ true)},$$

where the approximate degrees of freedom is given by

$$\nu \equiv \frac{\left(\frac{S_x^2}{n} + \frac{S_y^2}{m}\right)^2}{\frac{(S_x^2/n)^2}{n-1} + \frac{(S_y^2/m)^2}{m-1}}.$$

The following table summarizes how to carry out the various two- and one-sided tests.

two-sided	$H_0 : \mu_x = \mu_y$ $H_1 : \mu_x \neq \mu_y$	Reject H_0 iff $\lvert T_0^\star \rvert > t_{\alpha/2,\nu}$
one-sided	$H_0 : \mu_x \leq \mu_y$ $H_1 : \mu_x > \mu_y$	Reject H_0 iff $T_0^\star > t_{\alpha,\nu}$
one-sided	$H_0 : \mu_x \geq \mu_y$ $H_1 : \mu_x < \mu_y$	Reject H_0 iff $T_0^\star < -t_{\alpha,\nu}$

Example: Let's test $H_0 : \mu_x = \mu_y$ vs. $H_1 : \mu_x \neq \mu_y$ at level $\alpha = 0.10$.

Suppose we have the following data:

$$\begin{aligned} n &= 15, \quad \bar{X} = 24.2, \quad S_x^2 = 10 \\ m &= 10, \quad \bar{Y} = 23.9, \quad S_y^2 = 20. \end{aligned}$$

S_x^2 isn't very close to S_y^2, so we'll assume $\sigma_x^2 \neq \sigma_y^2$.

Plug-and-chug to get

$$T_0^\star = 0.184, \quad \nu = 14.9 \doteq 15, \quad \text{and} \quad t_{\alpha/2,\nu} = t_{0.05,15} = 1.753.$$

Since $|T_0^\star| < t_{\alpha/2,\nu}$, we *fail to reject* (i.e., grudgingly accept) H_0. Actually, since $T_0^\star$ was so close to 0, we didn't really need the tables. $\square$

Lesson 7.8 — Two-Sample Normal Means Test with Paired Observations

Again consider two competing normal populations. Suppose we collect observations from the two populations in *pairs*.

The RV's between *different* pairs are *independent*. The two observations within the *same* pair may *not* be independent — in fact, it's often good for them to be positively correlated (as explained in the previous module)!

Example: One twin takes a new drug, the other takes a placebo.

$$\text{independent} \begin{cases} \text{Pair } 1: & (X_1, Y_1) \\ \text{Pair } 2: & (X_2, Y_2) \\ \quad\vdots & \quad\vdots \\ \text{Pair } n: & \underbrace{(X_n, Y_n)}_{\text{not indep}} \end{cases}$$

Define the pair-wise differences,

$$D_i \;\equiv\; X_i - Y_i, \quad i = 1, 2, \ldots, n.$$

Note that $D_1, D_2, \ldots, D_n \stackrel{\text{iid}}{\sim} \text{Nor}(\mu_d, \sigma_d^2)$, where

$$\mu_d \;\equiv\; \mu_x - \mu_y \quad \text{and} \quad \sigma_d^2 \;\equiv\; \sigma_x^2 + \sigma_y^2 - 2\text{Cov}(X_i, Y_i).$$

Define the sample mean and variance of the differences,

$$\bar{D} \;\equiv\; \sum_{i=1}^{n} D_i/n \;\sim\; \text{Nor}(\mu_d, \sigma_d^2/n)$$

$$S_d^2 \;\equiv\; \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \bar{D})^2 \;\sim\; \frac{\sigma_d^2 \chi^2(n-1)}{n-1}.$$

Then the test statistic is (assuming $\mu_d = \mu_x - \mu_y = 0$)

$$T_0 \equiv \frac{\bar{D}}{\sqrt{S_d^2/n}} \sim t(n-1).$$

Using the exact same manipulations as in the single-sample normal mean problem with unknown variance, we get the following....

two-sided	$H_0 : \mu_d = 0$ $H_1 : \mu_d \neq 0$	Reject H_0 iff $\lvert T_0\rvert > t_{\alpha/2,n-1}$
one-sided	$H_0 : \mu_d \leq 0$ $H_1 : \mu_d > 0$	Reject H_0 iff $T_0 > t_{\alpha,n-1}$
one-sided	$H_0 : \mu_d \geq 0$ $H_1 : \mu_d < 0$	Reject H_0 iff $T_0 < -t_{\alpha,n-1}$

Example: Times for (the same) people to parallel park two cars.

Person	Park Honda	Park Cadillac	Difference
1	10	20	-10
2	25	40	-15
3	5	5	0
4	20	35	-15
5	15	20	-5

Let's test $H_0 : \mu_h = \mu_c$ at level $\alpha = 0.10$.

We see that $n = 5$, $\bar{D} = -9$, $S_d^2 = 42.5$. This gives $|T_0| = 3.087$.

Meanwhile, $t_{0.05,4} = 2.13$, so we *reject* H_0. We conclude that $\mu_h \neq \mu_c$ (and it's probably the case that Hondas are easier to park). $\square$

Lesson 7.9 — Normal Variance Test

What's Coming Up in the Next Few Lessons: We'll look at a variety of tests for parameters other than the mean.

- The variance σ^2 of a normal distribution (this lesson).
- The ratio of variances σ_x^2/σ_y^2 from two normals.
- The Bernoulli success parameter p.
- The difference of success parameters, $p_x - p_y$, from two Bernoullis.

Set-up for the Normal Variance Test:

Suppose $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} \text{Nor}(\mu, \sigma^2)$, where μ and σ^2 are *unknown*.

Consider the two-sided test (where you specify the hypothesized σ_0^2):

$$H_0 : \sigma^2 = \sigma_0^2 \quad \text{vs.} \quad H_1 : \sigma^2 \neq \sigma_0^2.$$

Recall (yet again) that the sample variance

$$S^2 \equiv \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \sim \frac{\sigma^2 \chi^2(n-1)}{n-1}.$$

We'll use the test statistic

$$\chi_0^2 \equiv \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1) \quad \text{(if } H_0 \text{ is true).}$$

The two-sided test rejects H_0 iff

$$\chi_0^2 < \chi^2_{1-\alpha/2,n-1} \quad \text{or} \quad \chi_0^2 > \chi^2_{\alpha/2,n-1}.$$

One-Sided Tests:

$$H_0 : \sigma^2 \le \sigma_0^2 \quad \text{vs.} \quad H_1 : \sigma^2 > \sigma_0^2.$$

$$\text{Reject } H_0 \quad \text{iff} \quad \chi_0^2 > \chi^2_{\alpha,n-1}.$$

$$H_0 : \sigma^2 \ge \sigma_0^2 \quad \text{vs.} \quad H_1 : \sigma^2 < \sigma_0^2.$$

$$\text{Reject } H_0 \quad \text{iff} \quad \chi_0^2 < \chi^2_{1-\alpha,n-1}.$$

Example: Suppose we want to test at level 0.05, whether or not the variance of a certain process is ≤ 0.02, specifically,

$$H_0 : \sigma^2 \leq 0.02 \quad \text{vs.} \quad H_1 : \sigma^2 > 0.02.$$

If the sample variance is "too high," we'll reject H_0.

Suppose we have $n = 20$, $\bar{X} = 125.12$, and $S^2 = 0.0225$.

Then the test statistic $\chi_0^2 = (n-1)S^2/\sigma_0^2 = 21.375$ (and isn't explicitly dependent on $\bar{X}$).

Further, $\chi^2_{\alpha,n-1} = \chi^2_{0.05,19} = 30.14$.

So we *fail to reject* H_0. $\square$

Lesson 7.10 — Two-Sample Normal Variances Test

Do the two populations have the same variance?

The usual set-up:

$$X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Nor}(\mu_x, \sigma_x^2)$$
$$Y_1, Y_2, \ldots, Y_m \overset{\text{iid}}{\sim} \text{Nor}(\mu_y, \sigma_y^2).$$

Assume all X's and Y's are independent.

We'll estimate the variances σ_x^2 and σ_y^2 by the sample variances S_x^2 and S_y^2.

Two-sided test: $H_0 : \sigma_x^2 = \sigma_y^2$ (or $H_0 : \sigma_x^2/\sigma_y^2 = 1$) vs. $H_1 : \sigma_x^2 \neq \sigma_y^2$.

We'll use the test statistic

$$F_0 \;\equiv\; \frac{S_x^2}{S_y^2} \;\sim\; F(n-1, m-1) \quad \text{(if } H_0 \text{ is true).}$$

Thus, we reject H_0 iff

$$F_0 \;<\; F_{1-\alpha/2,n-1,m-1} \quad \text{or} \quad F_0 \;>\; F_{\alpha/2,n-1,m-1}.$$

iff (because of a property of the F distribution discussed earlier)

$$F_0 \;<\; \frac{1}{F_{\alpha/2,m-1,n-1}} \quad \text{or} \quad F_0 \;>\; F_{\alpha/2,n-1,m-1}.$$

One-Sided Tests:

$$H_0: \sigma_x^2 \le \sigma_y^2 \quad \text{vs.} \quad H_1: \sigma_x^2 > \sigma_y^2.$$

$$\text{Reject } H_0 \quad \text{iff} \quad F_0 \; > \; F_{\alpha,n-1,m-1}.$$

$$H_0: \sigma_x^2 \ge \sigma_y^2 \quad \text{vs.} \quad H_1: \sigma_x^2 < \sigma_y^2.$$

$$\text{Reject } H_0 \quad \text{iff} \quad F_0 \; < \; F_{1-\alpha,n-1,m-1} \; = \; 1/F_{\alpha,m-1,n-1}.$$

Example: Suppose we want to test at level 0.05 whether or not two processes have the same variance.

$$H_0 : \sigma_x^2 = \sigma_y^2 \quad \text{vs.} \quad H_1 : \sigma_x^2 \neq \sigma_y^2.$$

If the ratio of the sample variances is "too high" or "too low," reject H_0.

Data: $n = 7$ observations with $S_x^2 = 7.78$; and $m = 8$ with $S_y^2 = 12.04$.

Then $F_0 = S_x^2/S_y^2 = 0.646$,

$F_{1-\alpha/2,n-1,m-1} = \frac{1}{F_{\alpha/2,m-1,n-1}} = \frac{1}{F_{0.025,7,6}} = 1/5.695 = 0.176$, and

$F_{\alpha/2,n-1,m-1} = F_{0.025,6,7} = 5.119$.

Since $F_{1-\alpha/2,n-1,m-1} \leq F_0 \leq F_{\alpha/2,n-1,m-1}$, we *fail to reject* H_0. $\square$

Lesson 7.11 — Bernoulli Proportion Test

Suppose that $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bern}(p)$.

We're interested in testing hypotheses about the success parameter p.

Two-sided test (you specify the hypothesized p_0):

$$H_0 : p = p_0 \quad \text{vs.} \quad H_1 : p \neq p_0.$$

Let $Y = \sum_{i=1}^n X_i \sim \text{Bin}(n, p)$.

We'll use the test statistic

$$Z_0 \;\equiv\; \frac{Y - np_0}{\sqrt{np_0(1-p_0)}} \;=\; \frac{\bar{X} - p_0}{\sqrt{p_0(1-p_0)/n}}.$$

If H_0 is true, the central limit theorem implies that

$$Z_0 \;\approx\; \text{Nor}(0, 1).$$

Georgia Tech

Thus, for the two-sided test, we reject H_0 iff $|Z_0| > z_{\alpha/2}$.

Remark: In order for the CLT to work, you need n large (say at least 30), and both $np \geq 5$ and $nq \geq 5$ (so that p isn't too close to 0 or 1).

Remark: If n isn't very big, you may have to use Binomial tables (instead of the normal approximation). This gets tedious, and I won't go into it here!

One-Sided Tests:

$$H_0 : p \leq p_0 \quad \text{vs.} \quad H_1 : p > p_0.$$

$$\text{Reject } H_0 \quad \text{iff} \quad Z_0 \ > \ z_\alpha.$$

$$H_0 : p \geq p_0 \quad \text{vs.} \quad H_1 : p < p_0.$$

$$\text{Reject } H_0 \quad \text{iff} \quad Z_0 \ < \ -z_\alpha.$$

Example: In 200 samples of a certain semiconductor, there were only 4 defectives. We're interested in proving "beyond a shadow of a doubt" that the probability of a defective is less than 0.06. Let's conduct the test at level 0.05.

$$H_0 : p \geq 0.06 \quad \text{vs.} \quad H_1 : p < 0.06.$$

(Since p is close to 0, we really did need to take a lot of observations — 200 in this case — in order for the CLT to work.)

We have $n = 200$, $Y = 4$ defectives, and $p_0 = 0.06$.

The test statistic is

$$Z_0 = \frac{Y - np_0}{\sqrt{np_0(1 - p_0)}} = -2.357.$$

Since $-z_\alpha = -1.645$, we *reject* H_0; so it seems p really is < 0.06. $\square$

Sample-Size Selection

Can we **design** a two-sided test $H_0 : p = p_0$ vs. $H_1 : p \neq p_0$ such that

$$P(\text{Type I error}) \;=\; \alpha \quad \text{and} \quad P(\text{Type II error} \,|\, p \neq p_0) \;=\; \beta?$$

Yes! We'll now show that the necessary sample size is

$$n \;\approx\; \left[\frac{z_{\alpha/2}\sqrt{p_0 q_0} + z_\beta \sqrt{pq}}{p - p_0}\right]^2,$$

where, to save space, we let $q \equiv 1 - p$ and $q_0 \equiv 1 - p_0$.

Note that n is a function of the unknown p. In practice, we'll choose some $p = p_1$ and ask "How many observations should I take if p happens to equal p_1 instead of p_0" (where you pick p_1)? Thus, we guard against the scenario in which p actually equals p_1.

Proof (similar to normal mean design proof):

$$
\begin{aligned}
\beta &= P(\text{Type II error}) \\
&= P(\text{Fail to Reject } H_0 \,|\, H_0 \text{ false}) \\
&\doteq P(|Z_0| \le z_{\alpha/2} \,|\, p \ne p_0) \quad \text{(by the CLT)} \\
&= P(-z_{\alpha/2} \le Z_0 \le z_{\alpha/2} \,|\, p \ne p_0) \\
&= P\left(-z_{\alpha/2} \le \frac{Y - np_0}{\sqrt{np_0(1-p_0)}} \le z_{\alpha/2} \,\middle|\, p \ne p_0\right) \\
&= P\left(-z_{\alpha/2}\sqrt{\frac{p_0 q_0}{pq}} \le \frac{Y - np_0}{\sqrt{npq}} \le z_{\alpha/2}\sqrt{\frac{p_0 q_0}{pq}} \,\middle|\, p \ne p_0\right) \\
&= P\left(-c \le \frac{Y - np}{\sqrt{npq}} + \frac{n(p - p_0)}{\sqrt{npq}} \le c \,\middle|\, p \ne p_0\right),
\end{aligned}
$$

where

$$
c \equiv z_{\alpha/2}\sqrt{\frac{p_0 q_0}{pq}}.
$$

Now notice that (since p is the true success probability),

$$Z \;\equiv\; \frac{Y - np}{\sqrt{npq}} \;\approx\; \text{Nor}(0,1).$$

This gives

$$\begin{aligned}
\beta \;&\doteq\; P\left(-c \;\leq\; Z + \frac{n(p-p_0)}{\sqrt{npq}} \;\leq\; c\right) \\
&=\; P\left(-c - \frac{\sqrt{n}(p-p_0)}{\sqrt{pq}} \;\leq\; Z \;\leq\; c - \frac{\sqrt{n}(p-p_0)}{\sqrt{pq}}\right) \\
&=\; P(-c - d \;\leq\; Z \;\leq\; c - d) \\
&=\; \Phi(c-d) - \Phi(-c-d),
\end{aligned}$$

where

$$d \;\equiv\; \frac{\sqrt{n}(p-p_0)}{\sqrt{pq}}.$$

Also notice that

$$-c-d \;=\; -z_{\alpha/2}\sqrt{\frac{p_0q_0}{pq}}-\frac{\sqrt{n}(p-p_0)}{\sqrt{pq}} \;\ll\; 0.$$

This implies $\Phi(-c-d) \doteq 0$, and so $\beta \;\doteq\; \Phi(c-d)$. Thus,

$$-z_\beta \;\equiv\; \Phi^{-1}(\beta) \;\doteq\; c-d \;=\; z_{\alpha/2}\sqrt{\frac{p_0q_0}{pq}}-\frac{\sqrt{n}(p-p_0)}{\sqrt{pq}}.$$

After a little algebra, we finally(!) get

$$n \;\doteq\; \left[\frac{z_{\alpha/2}\sqrt{p_0q_0}+z_\beta\sqrt{pq}}{p-p_0}\right]^2.$$

Similarly, the sample size for the corresponding one-sided test is

$$n \;\doteq\; \left[\frac{z_{\alpha}\sqrt{p_0q_0}+z_\beta\sqrt{pq}}{p-p_0}\right]^2. \quad \text{Whew!}$$

Example: We're conducting a study on whether or not a particular allergy medication works effectively. We'll assume that the drug either clearly works or doesn't work for each independent subject, so that we'll have legitimate Bernoulli trials. Our hypothesis is $H_0 : p = p_0 = 0.8$ vs. $H_1 : p \neq 0.8$.

In order to design our test (i.e., determine its sample size), let's set our Type I error probability to $\alpha = 0.05$.

We'd like to protect against goofing up on the poor performance side, so let's set our Type II to error to $\beta = 0.10$ in the special case that $p = p_1 = 0.7$. Then

$$\begin{aligned} n &\doteq \left[\frac{z_{\alpha/2}\sqrt{p_0 q_0} + z_\beta \sqrt{p_1 q_1}}{p_1 - p_0}\right]^2 \\ &= \left[\frac{1.96\sqrt{(0.8)(0.2)} + 1.28\sqrt{(0.7)(0.3)}}{0.7 - 0.8}\right]^2 \\ &= 187.8 \;\rightarrow\; 188. \quad \Box \end{aligned}$$

Georgia Tech

Lesson 7.12 — Two-Sample Bernoulli Proportions Test

Suppose that $X_1, X_2, \ldots, X_n \stackrel{\text{iid}}{\sim} \text{Bern}(p_x)$ and $Y_1, Y_2, \ldots, Y_m \stackrel{\text{iid}}{\sim} \text{Bern}(p_y)$ are two independent Bernoulli samples.

Now we're interested in testing hypotheses about the difference in the success parameters, $p_x - p_y$.

Two-sided test:

$$H_0 : p_x = p_y \quad \text{vs.} \quad H_1 : p_x \neq p_y.$$

Denote the respective sample means by

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \frac{\text{Bin}(n, p_x)}{n} \quad \text{and} \quad \bar{Y} \sim \frac{\text{Bin}(m, p_y)}{m}.$$

By the CLT and our confidence interval work from the previous module, we know that for large n and m,

$$\bar{X} \approx \text{Nor}\left(p_x, \frac{p_x(1-p_x)}{n}\right) \quad \text{and} \quad \bar{Y} \approx \text{Nor}\left(p_y, \frac{p_y(1-p_y)}{m}\right).$$

Then

$$\frac{\bar{X} - \bar{Y} - (p_x - p_y)}{\sqrt{\frac{p_x(1-p_x)}{n} + \frac{p_y(1-p_y)}{m}}} \approx \text{Nor}(0,1).$$

Moreover, under the null hypothesis, we have $p \equiv p_x = p_y$, in which case

$$\frac{\bar{X} - \bar{Y}}{\sqrt{p(1-p)\left[\frac{1}{n} + \frac{1}{m}\right]}} \approx \text{Nor}(0,1).$$

Of course, p in the above equation is unknown. But if H_0 is true, then $p = p_x = p_y$, so let's estimate p by the **pooled estimator,**

$$\hat{p} \equiv \frac{\sum_{i=1}^{n} X_i + \sum_{j=1}^{m} Y_j}{n+m}.$$

Now plug this into the previous equation to get one last approximation — the test statistic that we can finally work with:

$$Z_0 \equiv \frac{\bar{X} - \bar{Y}}{\sqrt{\hat{p}(1-\hat{p})\left[\frac{1}{n} + \frac{1}{m}\right]}} \approx \text{Nor}(0,1) \quad \text{(under } H_0\text{)}.$$

Thus, for the two-sided test, we reject H_0 iff $|Z_0| > z_{\alpha/2}$.

One-Sided Tests:

$H_0 : p_x \le p_y$ vs. $H_1 : p_x > p_y \Rightarrow$ reject H_0 iff $Z_0 > z_\alpha$.

$H_0 : p_x \ge p_y$ vs. $H_1 : p_x < p_y \Rightarrow$ reject H_0 iff $Z_0 < -z_\alpha$.

Example: Compare two restaurants based on customer reviews (either yummy or nasty). Burger Fil-A got 178 yummies out of $n = 260$ reviews (+ 82 nasties), while McWendy's got 250 yummies (+ 50 nasties) out of $m = 300$ reviews.

Test $H_0 : p_b = p_m$ vs. $H_1 : p_b \neq p_m$ at level $\alpha = 0.05$. We have

$$\bar{B} = \frac{178}{260} = 0.6846, \quad \bar{M} = \frac{250}{300} = 0.8333, \quad \text{and} \quad \hat{p} = \frac{178 + 250}{260 + 300} = 0.7643.$$

This gives us

$$Z_0 = \frac{\bar{B} - \bar{M}}{\sqrt{\hat{p}(1 - \hat{p})\left[\frac{1}{n} + \frac{1}{m}\right]}} = \frac{0.6846 - 0.8333}{\sqrt{0.7643(0.2357)\left[\frac{1}{260} + \frac{1}{300}\right]}} = -4.135.$$

Since $|Z_0| > z_{0.025} = 1.96$, we easily reject H_0, and informally declare McWendy's the winner. $\Box$

Lesson 7.13 — Goodness-of-Fit Tests: Introduction

At this point, let's suppose that we've guessed a reasonable distribution and then estimated the relevant parameters. Now let's conduct a formal test to see just how successful our toils have been — in other words, is our hypothesized distribution + relevant parameters acceptable?

In particular, we'll carry out a **goodness-of-fit test**,

$$H_0 : X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{pmf / pdf } f(x).$$

What's Coming Up:

- Introduction (this lesson)
- Various examples (next lesson)
- A tough honors example (final lesson)

High-level view of a goodness-of-fit test procedure:

1. Divide the domain of $f(x)$ into k sets, say, $A_1, A_2, \ldots, A_k$ (distinct points if X is discrete, or intervals if X is continuous).
2. Tally the actual number of observations O_i that fall in set A_i, $i = 1, 2, \ldots, k$. Define $p_i \equiv P(X \in A_i)$, so $O_i \sim \text{Bin}(n, p_i)$.
3. Determine the expected number of observations that would fall in each set if H_0 were true, say, $E_i = \text{E}[O_i] = np_i$, $i = 1, 2, \ldots, k$.
4. Calculate a test statistic based on the differences between the E_i's and O_i's. The **chi-squared g-o-f test** statistic is

$$\chi_0^2 \equiv \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}.$$

(Why does the above test statistic remind me of Old McDonald's Farm? ☺)

5. A large value of χ_0^2 indicates a bad fit (so just do one-sided test).

   We *reject* H_0 if $\chi_0^2 > \chi^2_{\alpha,k-1-s}$, where

   - s is the number of unknown parameters from $f(x)$ that have to be estimated. E.g., if $X \sim \text{Nor}(\mu, \sigma^2)$, then $s = 2$.
   - $\chi^2_{\alpha,\nu}$ is the $(1-\alpha)$ quantile of the $\chi^2(\nu)$ distribution, i.e., $P\big(\chi^2(\nu) < \chi^2_{\alpha,\nu}\big) = 1 - \alpha$.

   If $\chi_0^2 \le \chi^2_{\alpha,k-1-s}$, we *fail to reject* (grudgingly accept) H_0.

Remarks:

- Usual recommendation: For the χ^2 g-o-f test to work, pick k, n such that $E_i \geq 5$ for all i, and n is at least 30.

- If the df $\nu = k - 1 - s$ happens to be very big, then

$$\chi^2_{\alpha,\nu} \approx \nu \left[1 - \frac{2}{9\nu} + z_\alpha \sqrt{\frac{2}{9\nu}} \right]^3,$$

  where z_α is the appropriate standard normal quantile.

- Other g-o-f tests: Kolmogorov–Smirnov, Anderson–Darling, Shapiro–Wilk, etc.

Baby Example: Test H_0: X_i's are Unif(0,1), with $\alpha = 0.05$.

Suppose we have $n = 1000$ observations divided into $k = 5$ equal-length intervals.

interval	[0,0.2]	(0.2,0.4]	(0.4,0.6]	(0.6,0.8]	(0.8,1.0]
p_i	0.2	0.2	0.2	0.2	0.2
$E_i = np_i$	200	200	200	200	200
O_i	179	208	222	199	192

It turns out that $\chi_0^2 \equiv \sum_{i=1}^{k}(O_i - E_i)^2/E_i = 5.27$.

No unknown parameters, so $s = 0$. Then $\chi^2_{\alpha,k-1-s} = \chi^2_{0.05,4} = 9.49$.

Since $\chi_0^2 \leq \chi^2_{\alpha,k-1-s}$, we fail to reject H_0. So we'll grudgingly pretend that the numbers are uniform. $\square$

Lesson 7.14 — Goodness-of-Fit Tests: Examples

We'll do a couple of detailed examples in this lesson.

Discrete Example: The number of defects in printed circuit boards is hypothesized to follow a Geometric(p) distribution. We collect a random sample of $n = 70$ printed boards, and we observe the number of defects in each.

# defects	frequency
1	34
2	18
3	2
4	9
5	7
	70

Georgia Tech

Start by getting the maximum likelihood estimator for the geometric parameter p. The likelihood function is

$$
\begin{aligned}
L(p) &= \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} (1-p)^{x_i - 1} p = (1-p)^{\sum_{i=1}^{n} x_i - n} p^n \\
\ell\text{n}(L(p)) &= \Big(\sum_{i=1}^{n} x_i - n \Big) \ell\text{n}(1-p) + n\,\ell\text{n}(p) \\
\frac{d\,\ell\text{n}(L(p))}{dp} &= \frac{-\sum_{i=1}^{n} x_i + n}{1-p} + \frac{n}{p} = 0.
\end{aligned}
$$

Solving for p gives the MLE,

$$
\hat{p} = \frac{1}{\bar{X}} = \frac{70}{1(34) + 2(18) + 3(2) + 4(9) + 5(7)} = 0.4762.
$$

Georgia Tech

Let's get the g-o-f test statistic, χ_0^2. We'll make a little table, assuming $\hat{p} = 0.4762$ is correct. By the Invariance Property of MLEs (this is why we learned it!), the expected number of boards having a certain value x is $E_x = nP(X = x) = n(1 - \hat{p})^{x-1}\hat{p}$ (assuming $\hat{p}$ is actually p).

x	$P(X = x)$	E_x	O_x
1	0.4762	33.33	34
2	0.2494	17.46	18
3	0.1307	9.15	2
4	0.0684	4.79	9
$\geq 5^\star$	0.0753	5.27	7
	1.0000	70	70

*Combine the entries in the last row (≥ 5) so the probabilities sum to one.

Well, we really ought to combine the last two cells too, since $E_4 = 4.79 < 5$. Let's do so to get the following "improved" table.

x	$P(X = x)$	E_x	O_x
1	0.4762	33.33	34
2	0.2494	17.46	18
3	0.1307	9.15	2
≥ 4	0.1437	10.06	16
	1.0000	70	70

Thus, the test statistic is

$$\chi_0^2 \;=\; \sum_{x=1}^{4} \frac{(E_x - O_x)^2}{E_x} \;=\; \frac{(33.33 - 34)^2}{33.33} + \cdots \;=\; 9.12.$$

Let $k = 4$ denote the number of cells (that we ultimately ended up with), and let $s = 1$ denote the number of parameters we had to estimate.

Suppose the level $\alpha = 0.05$.

Then we compare $\chi_0^2 = 9.12$ against $\chi^2_{\alpha,k-1-s} = \chi^2_{0.05,2} = 5.99$.

Since $\chi_0^2 > \chi^2_{\alpha,k-1-s}$, we reject H_0.

This means that the number of defects probably isn't geometric. $\Box$

Continuous Distributions: For the continuous case, let's denote the intervals $A_i \equiv (a_{i-1}, a_i]$, $i = 1, 2, \ldots, k$. For convenience, we choose the a_i's to ensure that we have **equal-probability intervals**, i.e.,

$$p_i \;=\; P(X \in A_i) \;=\; P(a_{i-1} < X \leq a_i) \;=\; 1/k \quad \text{for all } i.$$

In this case, we immediately have $E_i = n/k$ for all i, and then

$$\chi_0^2 \;=\; \sum_{i=1}^{k} \frac{(O_i - (n/k))^2}{n/k}.$$

The issue is that the a_i's might depend on unknown parameters.

Example: Suppose that we're interested in fitting a distribution to a series of interarrival times. Could they be *Exponential*?

$$H_0 : X_1, X_2, \ldots, X_n \stackrel{\text{iid}}{\sim} \text{Exp}(\lambda).$$

Let's do a χ^2 g-o-f test with equal-probability intervals.

This amounts to choosing a_i's such that the cdf

$$F(a_i) \;=\; P(X \le a_i) \;=\; 1 - e^{-\lambda a_i} \;=\; \frac{i}{k}, \quad i = 0, 1, 2, \ldots, k.$$

That is, after a wee bit of algebra,

$$a_i \;=\; -\frac{1}{\lambda}\ell\text{n}\Big(1 - \frac{i}{k}\Big), \quad i = 0, 1, 2, \ldots, k.$$

Great, but λ is unknown (so we'll need to estimate $s = 1$ parameter).

Good News: We know that the MLE is $\widehat{\lambda} = 1/\bar{X}$. Thus, by the Invariance Property, the MLEs of the a_i's are

$$\widehat{a}_i \;=\; -\frac{1}{\widehat{\lambda}}\,\ell\text{n}\Big(1 - \frac{i}{k}\Big) \;=\; -\bar{X}\,\ell\text{n}\Big(1 - \frac{i}{k}\Big), \quad i = 0, 1, 2, \ldots, k.$$

Continue the Example: We take $n = 100$ observations and divide them into $k = 10$ equal-probability intervals, so that $E_i = n/k = 10$ for all i. Suppose that the sample mean based on the 100 observations is $\bar{X} = 0.8778$. Then

$$\widehat{a}_i \;=\; -0.8778\,\ell\text{n}\big(1 - 0.1i\big), \quad i = 0, 1, 2, \ldots, 10.$$

Further suppose we determine which interval each of the 100 observations belongs to and tally them up to get the O_i's. . . .

interval $(\hat{a}_{i-1}, \hat{a}_i]$	O_i	$E_i = n/k$
[0, 0.092]	0	10
(0.092, 0.196]	1	10
(0.196, 0.313]	1	10
(0.313, 0.448]	6	10
(0.448, 0.608]	17	10
(0.608, 0.804]	21	10
(0.804, 1.057]	23	10
(1.057, 1.413]	24	10
(1.413, 2.021]	7	10
(2.021, ∞)	0	10
	100	100

$$\chi_0^2 = \sum_{i=1}^{k} (O_i - E_i)^2 / E_i = 92.2 \quad \text{and} \quad \chi^2_{\alpha,k-1-s} = \chi^2_{0.05,8} = 15.51.$$

So reject H_0. These guys ain't Expo. No way, no how. ☹

Georgia Tech

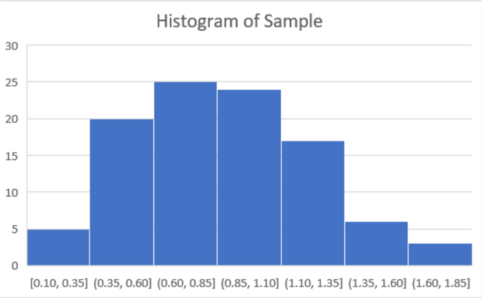
Lesson 7.15 — Goodness-of-Fit Tests: Honors Example

Let's make things more interesting with an extended example / mini-project.

We consider the same sample of 100 iid observations from the end of the last lesson, now with some more details.

$$0.9448 \qquad 0.8332 \qquad 0.6811 \qquad \cdots \qquad 0.5635$$

The sample mean for this data set is $\bar{X} = 0.8778$, and the sample standard deviation is $S = 0.3347$. Here's what the little fella looks like:



In this lesson, we'll do various goodness-of-fit tests to determine which distribution(s) the data could come from.

$$H_0 : X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} f(x),$$

where we'll consider the following possibilities:

- Exponential (rejected in the previous lesson).
- Gamma (which generalizes the exponential).
- Weibull (which generalizes the exponential in a different way).

In each case, we'll divide the data into $k = 10$ equal-probability intervals and perform a χ^2 g-o-f test at level $\alpha = 0.05$.

Along the way, we'll encounter a number of interesting issues that we'll need to deal with; but it'll all work out in the end. ☺

Exponential: In the previous lesson, we tested the null hypothesis that

$$H_0 : X_1, X_2, \ldots, X_n \stackrel{\text{iid}}{\sim} \text{Exp}(\lambda).$$

Recall that we failed miserably.

But, in retrospect, this makes sense in light of the facts that:

- The graph doesn't look anywhere near exponential.
- The expected value and standard deviation of an $\text{Exp}(\lambda)$ random variable are both $1/\lambda$; yet the sample mean $\bar{X} = 0.8778$ is $\gg$ the sample standard deviation $S = 0.3347$.

So this motivates our need to look at other distributions in our quest to find a good data fit.

Gamma: The Gamma distribution with parameters r and λ has pdf

$$f(x) \; = \; \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, \quad x > 0.$$

Note that $r = 1$ yields the Exp(λ) as a special case. We'll test

$$H_0 : X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Gam}(r, \lambda).$$

Way back in the good old days, we found the MLEs for r and λ:

$$\hat{\lambda} \; = \; \hat{r}/\bar{X},$$

where $\hat{r}$ solves

$$g(r) \; \equiv \; n\,\ell\text{n}(r/\bar{X}) - n\Psi(r) + \ell\text{n}\Big(\prod_{i=1}^{n} X_i\Big) \; = \; 0,$$

and where $\Psi(r) \equiv \Gamma'(r)/\Gamma(r)$ is the *digamma function*.

Since the digamma is sometimes a little hard to find in the usual software packages, we'll incorporate the approximation

$$\Gamma'(r) \doteq \frac{\Gamma(r+h)-\Gamma(r)}{h} \quad \text{(for any small } h \text{ of your choosing).}$$

So we need to find $\hat{r}$ that solves

$$g(r) \doteq n\,\ell\text{n}(r) - n\,\ell\text{n}(\bar{X}) - \frac{n}{h}\left(\frac{\Gamma(r+h)}{\Gamma(r)} - 1\right) + \ell\text{n}\left(\prod_{i=1}^{n} X_i\right) = 0. \quad (1)$$

How to solve for a zero?

- trial-and-error or by some sort of linear search — that's for losers!
- **bisection method** — let's try it here!
- **Newton's method** — stay tuned!

Bisection is an easy way to find a zero of any continuous function $g(r)$. It relies on the **Intermediate Value Theorem (IVT)**, which states that if $g(\ell)g(u) < 0$, then there is a zero $r^\star \in [\ell, u]$. Using this fact, it's easy to hone in on a zero via sequential bisecting:

- Initialization: Find lower and upper bounds $\ell_0 < u_0$ such that $g(\ell_0)g(u_0) < 0$. Then the IVT implies that $r^\star \in [\ell_0, u_0]$.
- For $i = 1, 2, \ldots$,
  - Let the midpoint of the current interval be $r_{i+1} \leftarrow (\ell_i + u_i)/2$.
  - If $g(r_{i+1})$ is sufficiently close to 0, or the interval width $u_i - \ell_i$ is sufficiently small, or your iteration budget is exceeded, then set $r^\star \leftarrow r_{i+1}$ and STOP.
  - If the sign of $g(r_{i+1})$ matches that of $g(\ell_i)$, this means that $r^\star \in [r_{i+1}, u_i]$; so set $\ell_{i+1} \leftarrow r_{i+1}$ and $u_{i+1} \leftarrow u_i$. Otherwise, $r^\star \in [\ell_i, r_{i+1}]$; so set $\ell_{i+1} \leftarrow \ell_i$ and $u_{i+1} \leftarrow r_{i+1}$.

Each iteration of the algorithm chops the search area in two and therefore converges to $r^\star$ pretty quickly.

Let's try bisection out on our dataset of $n = 100$ observations, where we recall that the sample mean is $\bar{x} = 0.8778$. And trust me that $\ell\text{n}\big(\prod_{i=1}^n x_i\big) = -21.5623$.

Let's take the approximate differentiation term $h = 0.01$. Then here's what Equation (1) simplifies to:

$$\begin{aligned} g(r) &\doteq 100\,\ell\text{n}(r) - 100\,\ell\text{n}(0.8778) - \frac{100}{0.01}\Big(\frac{\Gamma(r+0.01)}{\Gamma(r)} - 1\Big) - 21.5623 \\ &= 100\,\ell\text{n}(r) - \frac{10000\,\Gamma(r+0.01)}{\Gamma(r)} + 9991.47 \;=\; 0. \end{aligned}$$

In order to initialize the bisection algorithm, we note that $g(5) = 0.5506$ and $g(7) = -3.0595$. So there's a zero in there somewhere just itching to be found!

The algorithm is depicted in all of its glory in the next table.

step	ℓ_i	$g(\ell_i)$	u_i	$g(u_i)$	r_{i+1}	$g(r_{i+1})$
0	5.0000	0.5506	7.0000	-3.0595	6.0000	-1.5210
1	5.0000	0.5506	6.0000	-1.5210	5.5000	-0.5698
2	5.0000	0.5506	5.5000	-0.5698	5.2500	-0.0338
3	5.0000	0.5506	5.2500	-0.0338	5.1250	0.2519
4	5.1250	0.2519	5.2500	-0.0338	5.1875	0.1075
5	5.1875	0.1075	5.2500	-0.0338	5.2188	0.0365
6	5.2188	0.0365	5.2500	-0.0338	5.2344	0.0013
7	5.2344	0.0013	5.2500	-0.0338	5.2422	-0.0163
$\vdots$						
14	5.2349	0.0000	5.2349	0.0000	$\boldsymbol{r^\star = 5.2349}$	0.0000

We see that the algorithm eventually gives $\hat{r} = r^\star = 5.2349$; and then $\hat{\lambda} = \hat{r}/\bar{X} = 5.9637$.

So now we can start our χ^2 goodness-of-fit toils, noting that we have $s = 2$ unknown parameters.

We take the $n = 100$ observations and divide them into $k = 10$ equal-probability intervals, so that $E_i = n/k = 10$ for all i.

The (approximate) endpoints of the intervals are implicitly given by $\widehat{F}(\hat{a}_i) = i/k$, $i = 0, 1, 2, \ldots, k$, where $\widehat{F}(x)$ is the cdf of the Gam$(\hat{r}, \hat{\lambda})$ distribution.

Sadly, gamma distribution's cdf doesn't have a closed-form. But that's why we have Excel (or its friends) around, e.g.,

$$\hat{a}_i \;=\; \widehat{F}^{-1}(i/k) \;=\; \texttt{GAMMAINV}(i/k, \hat{r}, \hat{\lambda}).$$

interval $(\hat{a}_{i-1}, \hat{a}_i]$	O_i	$E_i = n/k$
[0.000, 0.436]	8	10
(0.436, 0.550]	12	10
(0.550, 0.644]	6	10
(0.644, 0.733]	9	10
(0.733, 0.823]	12	10
(0.823, 0.920]	8	10
(0.920, 1.032]	13	10
(1.032, 1.174]	14	10
(1.174, 1.391]	10	10
(1.391, ∞)	8	10
	100	100

$$\chi_0^2 = \sum_{i=1}^{k} (O_i - E_i)^2 / E_i = 6.2 \quad \text{and} \quad \chi^2_{\alpha,k-1-s} = \chi^2_{0.05,7} = 14.07.$$

So fail to reject H_0. These may indeed be gamma! ☺

Weibull: The Weibull distribution has cdf $F(x) = 1 - \exp\big[-(\lambda x)^r\big]$, for $x \geq 0$. Note that $r = 1$ yields the Exp(λ) as a special case.

Let's get MLEs for the $s = 2$ unknown parameters (r and λ). After a little algebra (a couple of chain rules), the pdf is

$$f(x) \;=\; \lambda r(\lambda x)^{r-1} e^{-(\lambda x)^r}, \quad x \geq 0.$$

So the likelihood function for an iid sample of size n is

$$L(r,\lambda) \;=\; \prod_{i=1}^{n} f(x_i) \;=\; \lambda^{nr} r^n \Big(\prod_{i=1}^{n} x_i\Big)^{r-1} \exp\Big[-\lambda^r \sum_{i=1}^{n} x_i^r\Big].$$

$$\ell\text{n}(L) \;=\; nr\,\ell\text{n}(\lambda) + n\,\ell\text{n}(r) + (r-1)\,\ell\text{n}\Big(\prod_{i=1}^{n} x_i\Big) - \lambda^r \sum_{i=1}^{n} x_i^r.$$

At this point, maximize with respect to r and λ by setting

$$\frac{\partial}{\partial r}\,\ell\mathrm{n}(L) \;=\; 0 \quad \text{and} \quad \frac{\partial}{\partial \lambda}\,\ell\mathrm{n}(L) \;=\; 0.$$

After more algebra — including the fact that $\frac{d}{dx}\,c^x = c^x \ell\mathrm{n}(c)$ — we get the simultaneous equations

$$\lambda \;=\; \Big(\frac{1}{n}\sum_{i=1}^{n} x_i^r\Big)^{-1/r} \quad \text{and}$$

$$g(r) \;=\; \frac{n}{r} + \ell\mathrm{n}\Big(\prod_{i=1}^{n} x_i\Big) - \frac{n\sum_i x_i^r\,\ell\mathrm{n}(x_i)}{\sum_i x_i^r} \;=\; 0.$$

The equation for λ looks easy enough, if only we could solve for r! ☹

But we can! Let's use **Newton's method**. It's usually a lot faster than bisection. Here's a reasonable implementation of Newton.

1. Initialize $r_0 = \bar{X}/S$, where $\bar{X}$ is the sample mean and S^2 is the sample variance. Set $j \leftarrow 0$.
2. Update $r_{j+1} \leftarrow r_j - g(r_j)/g'(r_j)$.
3. If $|g(r_{j+1})|$ or $|r_{j+1} - r_j|$ or your budget is suitably small, then STOP and set the MLE $\hat{r} \leftarrow r_{j+1}$. Otherwise, let $j \leftarrow j+1$ and goto Step 2.

To use Newton, we need (after yet more algebra)

$$g'(r) \;=\; -\frac{n}{r^2} - \frac{n\sum_i x_i^r\,[\ell\mathrm{n}(x_i)]^2}{\sum_i x_i^r} + \frac{n\big[\sum_i x_i^r\,\ell\mathrm{n}(x_i)\big]^2}{\big[\sum_i x_i^r\big]^2}.$$

Let's try Newton on our dataset of $n = 100$ observations, where $r_0 = \bar{X}/S = 0.8778/0.3347 = 2.6227$. This results in....

step	r_i	$g(r_i)$	$g'(r_i)$
0	2.6227	5.0896	−25.0848
1	2.8224	0.7748	−23.8654
2	2.8549	0.1170	−23.6493
3	2.8598	0.0178	−23.6174
4	2.8606	0.0027	−23.6126
5	2.8607		

Hence, $\hat{r} = r_5 = 2.8607$, and thus,

$$\widehat{\lambda} \;=\; \Big(\frac{1}{n}\sum_{i=1}^{n} x_i^{\hat{r}}\Big)^{-1/\hat{r}} \;=\; 1.0148.$$

Again do a χ^2 g-o-f test with equal-probability intervals. To get the endpoints, we note that $F(a_i) = i/k$; and then some algebra fun + the MLE Invariance Property yield

$$\begin{aligned} \hat{a}_i &= \frac{1}{\widehat{\lambda}}\left[-\ell\text{n}\left(1-\frac{i}{k}\right)\right]^{1/\hat{r}} \\ &= 0.9854\,[-\ell\text{n}(1-0.1\,i)]^{0.3496}, \quad i = 0,1,2,\ldots,10. \end{aligned}$$

Moreover, it turns out (see the next table) that

$$\chi_0^2 \;=\; \sum_{i=1}^{k}(O_i - E_i)^2/E_i \;=\; 5.0 \quad \text{and} \quad \chi^2_{\alpha,k-1-s} \;=\; \chi^2_{0.05,7} \;=\; 14.07.$$

So we fail to reject H_0, and we'll grudgingly pretend that these observations are Weibull. $\square$

interval $(\hat{a}_{i-1}, \hat{a}_i]$	O_i	$E_i = n/k$
[0, 0.4487]	8	10
(0.4487, 0.5833]	15	10
(0.5833, 0.6872]	9	10
(0.6872, 0.7792]	11	10
(0.7792, 0.8669]	8	10
(0.8669, 0.9557]	7	10
(0.9557, 1.0514]	10	10
(1.0514, 1.1637]	12	10
(1.1637, 1.3189]	11	10
(1.3189, ∞]	9	10
	100	100

The Big Reveal: I actually generated the observations from a Weibull distribution with parameters $r = 3$ and $\lambda = 1$. So we did pretty well! ☺