Hypothesis Testing for Proportions and Poisson

Hypothesis Testing Framework:

The first four steps don’t require using the data.

1) Specify Ho (null hypothesis).

2) Specify Ha (alternative hypothesis).

3) Decide on TS (test statistic). Often takes the form:
   \[
   \frac{\text{point estimate} - \text{expected value under null}}{\text{standard error of the point estimate}}
   \]

4) Determine RR (rejection region based on \(\alpha\))

The next # steps require the data.

5) Calculate the observed TS.

7) Determine if the observed TS falls in the RR and calculate the p-value appropriate for Ha.

8) Interpret the results in plain English for the context of the problem, including an appropriate confidence interval to convey a sense of magnitude and uncertainty.
**Hypothesis Testing for Proportions and Poisson**

\[ X \sim^{\text{approx}} \text{Normal Case} \]

Sample: \( x_1, x_2, \ldots, x_n \) iid \( F(\mu, \sigma^2) \) where \( F \) is close enough to a normal distribution that for the size of \( n \), the distribution of \( \bar{X} \) is essentially normal.

1) \( H_0: \mu = \mu_0 \).  
2) \( H_a: \mu \neq \mu_0 \). (two-sided alternative)  
3) \( TS = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \)  
4) \( RR: |TS| > t_{n-1, 0.975} \) for \( \alpha = 0.05 \).

Notice that we estimate the standard error of the point estimate from the data.

R command: `t.test`, specify null with option `mu = \mu_0`.
**X ~ Bernoulli Case**

Sample: $x_1, x_2, ..., x_n$ iid Bern($\theta$).

1) Ho: $\theta = \theta_0$.

2) Ha: $\theta \neq \theta_0$. (two-sided alternative)

3) $\text{TS} = \frac{\text{point estimate} - \text{expected value under null}}{\text{standard error of the point estimate}}$

We could use $\text{TS} = \frac{\hat{\theta} - \theta_0}{\sqrt{\frac{\theta(1-\theta)}{n}}}$

4) RR: $|\text{TS}| > Z_{0.975} = 1.96$ for $\alpha = 0.05$.

But this is silly*. We’re estimating the standard error of the point estimate when we’ve assumed it under the null.

* By “silly”, I mean anti-conservative, especially when $\theta_0$ is close to 0.5. When $\theta_0 = 0.5$, the estimated SE is highly likely to be < the SE under $H_0$. 

---

Author: Greevy, Blume  
BIOS 311  
Page 3 of 12
Better to use:

3) \( TS = \frac{\hat{\theta} - \theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}} \)

4) RR: \(|TS| > Z_{0.975} = 1.96\) for \(\alpha = 0.05\) and a two-sided alternative.

R command: prop.test, specify null with option \(p = \theta_0\). It defaults to 0.5.
Example: A hospital is studying how well health care practitioners adhere to hand washing protocols. Their current goal is to achieve > 90% compliance. In a sample of 600 appropriate hand washing situations, they observed practitioners washed their hands 552 times or 92% of the time. Is there sufficient evidence to say they are achieving their goal?

Set up the TS in your notes (Let’s use a two-sided alternative as opposed to the one-sided as the problem’s wording suggests.):
Hypothesis Testing for Proportions and Poisson

\( \text{Ho: } \theta = 0.90 \)

\( \text{Ha: } \theta \neq 0.90 \)

\[
TS = \frac{\hat{\theta} - 0.90}{\sqrt{\frac{0.90(1-0.90)}{600}}}
\]

RR: \( |TS| > Z_{0.975} = 1.96 \) for \( \alpha = 0.05 \) and a two-sided alternative.

\[
TS_{\text{obs}} = \frac{552/600-0.90}{\sqrt{\frac{0.90(1-0.90)}{600}}}
\]

\( = 1.6330. \)

\[
p\text{-value} = 2\text{pnorm}(-(552/600-0.9)/\text{sqrt}(.9*.1/600))
\]

\( = 0.1025 \)

Finish with a conclusion and nice CI, e.g.

We do not have sufficient evidence to conclude the true rate of hand washing is greater than 0.9 at a 5% significance level. The estimated rate was 0.92 with 95% confidence interval (0.8955, 0.9391).

library(Hmisc)
round( binconf(552,600), 4 )
The solution using R’s default test looks like this.

```r
> prop.test(552, 600, p = 0.90)

  1-sample proportions test with continuity correction
data:  552 out of 600, null probability 0.9
X-squared = 2.4491, df = 1, p-value = 0.1176
alternative hypothesis: true p is not equal to 0.9
95 percent confidence interval:
  0.8945962 0.9398571
sample estimates:
  p
  0.92
```

Asides: Chi-square df=1 and Continuity correction.

What assumption are we making in step 4?
2*pnorm(-(552/600-0.9)/sqrt(.9*.1/600))
vs
2*pnorm(-(551.5/600-0.9)/sqrt(.9*.1/600))

4) RR: |TS| > Z_{0.975} = 1.96 for α = 0.05.
Hypothesis Testing for Proportions and Poisson

What should we do if we don’t trust that assumption, or better, if simulations show that assumption is not reliable for our circumstances?

**Fisher’s Exact Test** (one sample case)

Recall the “exact” confidence interval we created for a binomial proportion. The “exact” test is based on the same idea, but is actually simpler to calculate. Just think carefully about the definition of a p-value.

\[ p\text{-value} = \text{the probability under } H_0 \text{ of observing something as extreme or more extreme than what you actually observed.} \]

In our example the expected number of hand washings was \( 0.9 \times 600 = 540 \). We observed 552. What’s as or more extreme than that?

552, 553, 554, ..., 600 are clearly as or more extreme.

If instead of 552, we had observed 528, then more extreme would have been 528, 527, 526, ..., 0. How come?
Rather than trying to figure out an appropriate opposite tail for what we actually observed, Rosner recommends taking the one-sided p-value for the observed tail and doubling it to get the two-sided p-value. (Since p-values can’t be greater than 1, round down to 1 if necessary.)

The solution in R, which uses a fancier approach to get the two-sided p-value, looks like this.

```r
> binom.test(552, 600, p = 0.90)

    Exact binomial test

data:  552 and 600
number of successes = 552, number of trials = 600,
p-value = 0.1170
alternative hypothesis: true probability of success is not equal to 0.9
95 percent confidence interval:  0.8953293 0.9404275
sample estimates:
probability of success  0.92
```

Why were the results so similar in this particular example?
Hypothesis Testing for Proportions and Poisson

Mirroring the large sample test and exact test for a Bernoulli, create the large sample and two-sample tests for the Poisson. I’ll start the solution.

**X ~ Poisson Case**

Sample: \( x_1, x_2, ..., x_n \) iid Poisson(\( \lambda \)) and \( n \) is large.

To simplify, let’s use that sum \( X_i s ~ Poisson(n\lambda) \) and make our inferences about sum \( X ~ Poisson(\theta) \), where \( \theta = n\lambda \).

1) Ho: \( \theta = \theta_o \). (Thus \( \theta_o = n\lambda_o \).)

2) Ha: \( \theta \neq \theta_o \). (two-sided alternative)

3) point estimate - expected value under null
   \[TS = \frac{\text{sum of } X_i - \theta_o}{\text{standard error of the point estimate}}\]

Now don’t go on auto-pilot here. What is the appropriate SE for our point estimate?
Hypothesis Testing for Proportions and Poisson

Finish specifying the TS and RR for the large sample case.

Specify how to get the p-value for the small sample case.
Two-sample proportions:  
\( X \sim \text{Bernoulli} \) and \( Y \sim \text{Bernoulli} \)

How is the null hypothesis different in how it gives us information about the SE in the two sample case?

It’s a little trickier to set this up in R.
\[
x \leftarrow \text{matrix}(c(552, 600-552, 950, 1000-950), \text{ncol} = 2)
x\text{chisq.test}(x)
\]