Introduction Based on a Simple Example*

STA442/2101 Fall 2012

Background Reading Optional

- Chapter 1 of *Data analysis with SAS*: What's going on and how would you say it to a client?
- Chapter 1 of Davison's *Statistical models*: Data, and probability models for data.

Steps in the process of statistical analysis One possible approach

- Consider a fairly realistic example or problem
- Decide on a statistical model
- Perhaps decide sample size
- Acquire data
- Examine and clean the data; generate displays and descriptive statistics
- Estimate parameters, perhaps by maximum likelihood
- Carry out tests, compute confidence intervals, or both
- Perhaps re-consider the model and go back to estimation
- Based on the results of estimation and inference, draw conclusions about the example or problem

Coffee taste test

A fast food chain is considering a change in the blend of coffee beans they use to make their coffee. To determine whether their customers prefer the new blend, the company plans to select a random sample of $n = 100$ coffee-drinking customers and ask them to taste coffee made with the new blend and with the old blend, in cups marked "A" and "B." Half the time the new blend will be in cup A, and half the time it will be in cup B. Management wants to know if there is a difference in preference for the two blends.

*See last slide for copyright information.

Statistical model

Letting θ denote the probability that a consumer will choose the new blend, treat the data $Y_1, \ldots, Y_n$ as a random sample from a Bernoulli distribution. That is, independently for $i = 1, \ldots, n$,

$$P(y_i|\theta) = \theta^{y_i}(1-\theta)^{1-y_i}$$

for $y_i = 0$ or $y_i = 1$, and zero otherwise.

Note that $Y = \sum_{i=1}^n Y_i$ is the number of consumers who choose the new blend. Because $Y \sim B(n, \theta)$, the whole experiment could also be treated as a single observation from a Binomial.

Find the MLE of θ Show your work

Denoting the likelihood by $L(\theta)$ and the log likelihood by $\ell(\theta) = \log L(\theta)$, maximize the log likelihood.

$$\begin{aligned}
\frac{\partial \ell}{\partial \theta} &= \frac{\partial}{\partial \theta} \log \left(\prod_{i=1}^n P(y_i|\theta) \right) \\
&= \frac{\partial}{\partial \theta} \log \left(\prod_{i=1}^n \theta^{y_i}(1-\theta)^{1-y_i} \right) \\
&= \frac{\partial}{\partial \theta} \log \left(\theta^{\sum_{i=1}^n y_i}(1-\theta)^{n-\sum_{i=1}^n y_i} \right) \\
&= \frac{\partial}{\partial \theta} \left((\sum_{i=1}^n y_i) \log \theta + (n - \sum_{i=1}^n y_i) \log(1-\theta) \right) \\
&= \frac{\sum_{i=1}^n y_i}{\theta} - \frac{n - \sum_{i=1}^n y_i}{1-\theta}
\end{aligned}$$

Setting the derivative to zero and solving

- $\theta = \frac{\sum_{i=1}^n y_i}{n} = \overline{y} = p$
- Second derivative test: $\frac{\partial^2 \log \ell}{\partial \theta^2} = -n \left(\frac{1-\overline{y}}{(1-\theta)^2} + \frac{\overline{y}}{\theta^2} \right) < 0$
- Concave down, maximum, and the MLE is the sample proportion.

Numerical estimate

Suppose 60 of the 100 consumers prefer the new blend. Give a point estimate the parameter θ. Your answer is a number.

```
> p = 60/100; p
[1] 0.6
```

Carry out a test to answer the question Is there a difference in preference for the two blends?

Start by stating the null hypothesis

- $H_0 : \theta = 0.50$
- $H_1 : \theta \neq 0.50$
- A case could be made for a one-sided test, but we'll stick with two-sided.
- $\alpha = 0.05$ as usual.
- Central Limit Theorem says $\widehat{\theta} = \overline{Y}$ is approximately normal with mean θ and variance $\frac{\theta(1-\theta)}{n}$.

Several valid test statistics for $H_0 : \theta = \theta_0$ are available Two of them are

$$Z_1 = \frac{\sqrt{n}(\overline{Y} - \theta_0)}{\sqrt{\theta_0(1-\theta_0)}}$$

and

$$Z_2 = \frac{\sqrt{n}(\overline{Y} - \theta_0)}{\sqrt{\overline{Y}(1-\overline{Y})}}$$

What is the critical value? Your answer is a number.

```
> alpha = 0.05
> qnorm(1-alpha/2)
[1] 1.959964
```

Calculate the test statistic and the p-value for each test Note: The R code uses p for the sample proportion

```
> theta0 = .5; p = .6; n = 100
> Z1 = sqrt(n)*(p-theta0)/sqrt(theta0*(1-theta0)); Z1
[1] 2
> pval1 = 2 * (1-pnorm(Z1)); pval1
[1] 0.04550026
>
```

```
> Z2 = sqrt(n)*(p-theta0)/sqrt(p*(1-p)); Z2
[1] 2.041241
> pval2 = 2 * (1-pnorm(Z2)); pval2
[1] 0.04122683
```

Conclusions

- Do you reject H_0? *Yes, just barely.*
- Isn't the $\alpha = 0.05$ significance level pretty arbitrary? *Yes, but if people insist on a Yes or No answer, this is what you give them.*
- What do you conclude, in symbols? $\theta \neq 0.50$. *Specifically,* $\theta > 0.50$.
- What do you conclude, in plain language? Your answer is a statement about coffee. *More consumers prefer the new blend of coffee beans.*
- Can you really draw directional conclusions when all you did was reject a non-directional null hypothesis? *Yes. Decompose the two-sided size* α *test into two one-sided tests of size* $\alpha/2$. *This approach works in general.*

It is very important to state directional conclusions, and state them clearly in terms of the subject matter. **Say what happened!** If you are asked state the conclusion in plain language, your answer *must* be free of statistical mumbo-jumbo.

What about negative conclusions? What would you say if $Z = 1.84$?

Here are two possibilities.

- "By conventional standards, this study does not provide enough evidence to conclude that consumers prefer one blend of coffee beans over the other."
- "The results are consistent with no difference in preference for the two coffee bean blends."

In this course, we will not just casually *accept* the null hypothesis.

Confidence Intervals Approximately for large n,

$$\begin{aligned}
1-\alpha &\approx Pr\{-z_{\alpha/2} < Z < z_{\alpha/2}\} \\
&= Pr\left\{-z_{\alpha/2} < \frac{\sqrt{n}(\overline{Y}-\theta)}{\sqrt{\overline{Y}(1-\overline{Y})}} < z_{\alpha/2}\right\} \\
&= Pr\left\{\overline{Y} - z_{\alpha/2}\sqrt{\frac{\overline{Y}(1-\overline{Y})}{n}} < \theta < \overline{Y} + z_{\alpha/2}\sqrt{\frac{\overline{Y}(1-\overline{Y})}{n}}\right\}
\end{aligned}$$

- Could express this as $\overline{Y} \pm z_{\alpha/2}\sqrt{\frac{\overline{Y}(1-\overline{Y})}{n}}$
- $z_{\alpha/2}\sqrt{\frac{\overline{Y}(1-\overline{Y})}{n}}$ is sometimes called the *margin of error*.
- If $\alpha = 0.05$, it's the 95% margin of error.

Give a 95% confidence interval for the taste test data. The answer is a pair of numbers. Show some work.

$$\begin{aligned}
& \left(\overline{y} - z_{\alpha/2}\sqrt{\frac{\overline{y}(1-\overline{y})}{n}} \quad , \quad \overline{y} + z_{\alpha/2}\sqrt{\frac{\overline{y}(1-\overline{y})}{n}}\right) \\
= & \left(0.60 - 1.96\sqrt{\frac{0.6 \times 0.4}{100}} \quad , \quad 0.60 + 1.96\sqrt{\frac{0.6 \times 0.4}{100}}\right) \\
= & (0.504, 0.696)
\end{aligned}$$

In a report, you could say

- The estimated proportion preferring the new coffee bean blend is 0.60 ± 0.096, or
- "Sixty percent of consumers preferred the new blend. These results are expected to be accurate within 10 percentage points, 19 times out of 20."

Meaning of the confidence interval

- We calculated a 95% confidence interval of $(0.504, 0.696)$ for θ.
- Does this mean $Pr\{0.504 < \theta < 0.696\} = 0.95$?
- No! The quantities 0.504, 0.696 and θ are all constants, so $Pr\{0.504 < \theta < 0.696\}$ is either zero or one.
- The endpoints of the confidence interval are random variables, and the numbers 0.504 and 0.696 are *realizations* of those random variables, arising from a particular random sample.
- Meaning of the probability statement: If we were to calculate an interval in this manner for a large number of random samples, the interval would contain the true parameter around 95% of the time.
- So we sometimes say that we are "95% confident" that $0.504 < \theta < 0.696$.

Confidence intervals (regions) correspond to tests Recall $Z_1 = \frac{\sqrt{n}(\overline{Y}-\theta_0)}{\sqrt{\theta_0(1-\theta_0)}}$ and $Z_2 = \frac{\sqrt{n}(\overline{Y}-\theta_0)}{\sqrt{\overline{Y}(1-\overline{Y})}}$.

From the derivation of the confidence interval,

$$-z_{\alpha/2} < Z_2 < z_{\alpha/2}$$

if and only if

$$\overline{Y} - z_{\alpha/2}\sqrt{\frac{\overline{Y}(1-\overline{Y})}{n}} < \theta_0 < \overline{Y} + z_{\alpha/2}\sqrt{\frac{\overline{Y}(1-\overline{Y})}{n}}$$

- So the confidence interval consists of those parameter values θ_0 for which $H_0 : \theta = \theta_0$ is *not* rejected.
- That is, the null hypothesis is rejected at significance level α if and only if the value given by the null hypothesis is outside the $(1-\alpha) \times 100\%$ confidence interval.
- There is a confidence interval corresponding to Z_1 too.
- In general, any test can be inverted to obtain a confidence region.

Selecting sample size

- Where did that $n = 100$ come from?
- Probably off the top of someone's head.
- We can (and should) be more systematic.
- Sample size can be selected
  - To achieve a desired margin of error
  - To achieve a desired statistical power
  - In other reasonable ways

Power

The power of a test is the probability of rejecting H_0 when H_0 is false.

- More power is good.
- Power is not just one number. It is a *function* of the parameter(s).
- Usually,
  - For any n, the more incorrect H_0 is, the greater the power.
  - For any parameter value satisfying the alternative hypothesis, the larger n is, the greater the power.

Statistical power analysis To select sample size

- Pick an effect you'd like to be able to detect – a parameter value such that H_0 is false. It should be just over the boundary of interesting and meaningful.
- Pick a desired power, a probability with which you'd like to be able to detect the effect by rejecting the null hypothesis.
- Start with a fairly small n and calculate the power. Increase the sample size until the desired power is reached.

There are two main issues.

- What is an "interesting" or "meaningful" parameter value?
- How do you calculate the probability of rejecting H_0?

Calculating power for the test of a single proportion True parameter value is θ

$$\begin{aligned}
\text{Power} &\approx 1 - Pr\{-z_{\alpha/2} < Z_2 < z_{\alpha/2}\} \\
&= 1 - Pr\left\{-z_{\alpha/2} < \frac{\sqrt{n}(\overline{Y}-\theta_0)}{\sqrt{\overline{Y}(1-\overline{Y})}} < z_{\alpha/2}\right\} \\
&= \ldots \\
&= 1 - Pr\left\{\frac{\sqrt{n}(\theta_0-\theta)}{\sqrt{\theta(1-\theta)}} - z_{\alpha/2}\sqrt{\frac{\overline{Y}(1-\overline{Y})}{\theta(1-\theta)}} < \frac{\sqrt{n}(\overline{Y}-\theta)}{\sqrt{\theta(1-\theta)}} \right. \\
&\qquad \left. < \frac{\sqrt{n}(\theta_0-\theta)}{\sqrt{\theta(1-\theta)}} + z_{\alpha/2}\sqrt{\frac{\overline{Y}(1-\overline{Y})}{\theta(1-\theta)}}\right\} \\
&\approx 1 - Pr\left\{\frac{\sqrt{n}(\theta_0-\theta)}{\sqrt{\theta(1-\theta)}} - z_{\alpha/2} < Z < \frac{\sqrt{n}(\theta_0-\theta)}{\sqrt{\theta(1-\theta)}} + z_{\alpha/2}\right\} \\
&= 1 - \Phi\left(\frac{\sqrt{n}(\theta_0-\theta)}{\sqrt{\theta(1-\theta)}} + z_{\alpha/2}\right) + \Phi\left(\frac{\sqrt{n}(\theta_0-\theta)}{\sqrt{\theta(1-\theta)}} - z_{\alpha/2}\right),
\end{aligned}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal.

An R function to calculate approximate power For the test of a single proportion

$$\text{Power} = 1 - \Phi\left(\frac{\sqrt{n}(\theta_0-\theta)}{\sqrt{\theta(1-\theta)}} + z_{\alpha/2}\right) + \Phi\left(\frac{\sqrt{n}(\theta_0-\theta)}{\sqrt{\theta(1-\theta)}} - z_{\alpha/2}\right)$$

```
Z2power = function(theta,n,theta0=0.50,alpha=0.05)
    {
    effect = sqrt(n)*(theta0-theta)/sqrt(theta*(1-theta))
    z = qnorm(1-alpha/2)
    Z2power = 1 - pnorm(effect+z) + pnorm(effect-z)
    Z2power
    } # End of function Z2power
```

Some numerical examples

```
> Z2power(0.50,100) # Should be alpha = 0.05
[1] 0.05
>
> Z2power(0.55,100)
[1] 0.1713209
> Z2power(0.60,100)
[1] 0.5324209
> Z2power(0.65,100)
[1] 0.8819698
> Z2power(0.40,100)
[1] 0.5324209
> Z2power(0.55,500)
[1] 0.613098
> Z2power(0.55,1000)
[1] 0.8884346
```

Find smallest sample size needed to detect $\theta = 0.60$ as different from $\theta_0 = 0.50$ with probability at least 0.80

```
> samplesize = 1
> power=Z2power(theta=0.60,n=samplesize); power
[1] 0.05478667
> while(power < 0.80)
+ {
+ samplesize = samplesize+1
+ power = Z2power(theta=0.60,n=samplesize)
+ }
> samplesize
[1] 189
> power
[1] 0.8013024
```

What is required of the scientist Who wants to select sample size by power analysis

The scientist must specify

- Parameter values that he or she wants to be able to detect as different from H_0 value.
- Desired power (probability of detection)

It's not always easy for a scientist to think in terms of the parameters of a statistical model.

Using the non-central chi-squared distribution For power and sample size calculations

If $X \sim N(\mu, \sigma^2)$, then

- $Z = \left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi^2(1)$
- $Y = \frac{X^2}{\sigma^2}$ is said to have a *non-central chi-squared* distribution with degrees of freedom one and *non-centrality parameter* $\lambda = \frac{\mu^2}{\sigma^2}$.
- Write $Y \sim \chi^2(1, \lambda)$

Facts about the non-central chi-squared distribution With one *df*

$$Y \sim \chi^2(1, \lambda), \text{ where } \lambda \geq 0$$

- $Pr\{Y > 0\} = 1$, of course.
- If $\lambda = 0$, the non-central chi-squared reduces to the ordinary central chi-squared.
- The distribution is "stochastically increasing" in λ, meaning that if $Y_1 \sim \chi^2(1, \lambda_1)$ and $Y_2 \sim \chi^2(1, \lambda_2)$ with $\lambda_1 > \lambda_2$, then $Pr\{Y_1 > y\} > Pr\{Y_2 > y\}$ for any $y > 0$.
- $\lim_{\lambda \to \infty} Pr\{Y > y\} = 1$
- There are efficient algorithms for calculating non-central chi-squared probabilities. R's `pchisq` function does it.

An example Back to the coffee taste test

$Y_1, \ldots, Y_n \stackrel{i.i.d.}{\sim} B(1, \theta)$

$H_0 : \theta = \theta_0 = \frac{1}{2}$

Reject H_0 if $|Z_2| = \left| \frac{\sqrt{n}(\overline{Y}-\theta_0)}{\sqrt{\overline{Y}(1-\overline{Y})}} \right| > z_{\alpha/2}$

Suppose that in the population, 60% of consumers would prefer the new blend. If we test 100 consumers, what is the probability of obtaining results that are statistically significant?

That is, if $\theta = 0.60$, what is the power with $n = 100$? Earlier, got 0.53 with a direct standard normal calculation.

Recall that if $X \sim N(\mu, \sigma^2)$, then $\frac{X^2}{\sigma^2} \sim \chi^2(1, \frac{\mu^2}{\sigma^2})$.

Reject H_0 if

$$|Z_2| = \left| \frac{\sqrt{n}(\overline{Y}-\theta_0)}{\sqrt{\overline{Y}(1-\overline{Y})}} \right| > z_{\alpha/2} \Leftrightarrow Z_2^2 > z_{\alpha/2}^2 = \chi^2_\alpha(1)$$

For large n, $X = \overline{Y} - \theta_0$ is approximately normal, with $\mu = \theta - \theta_0$ and $\sigma^2 = \frac{\theta(1-\theta)}{n}$. So,

$$\begin{aligned} Z_2^2 &= \frac{(\overline{Y}-\theta_0)^2}{\overline{Y}(1-\overline{Y})/n} \approx \frac{(\overline{Y}-\theta_0)^2}{\theta(1-\theta)/n} = \frac{X^2}{\sigma^2} \\ &\stackrel{approx}{\sim} \chi^2\left(1, n\,\frac{(\theta-\theta_0)^2}{\theta(1-\theta)}\right) \end{aligned}$$

We have found that

The Wald chi-squared test statistic of $H_0 : \theta = \theta_0$

$$Z_2^2 = \frac{n(\overline{Y}-\theta_0)^2}{\overline{Y}(1-\overline{Y})}$$

has an asymptotic non-central chi-squared distribution with $df = 1$ and non-centrality parameter

$$\lambda = n\,\frac{(\theta-\theta_0)^2}{\theta(1-\theta)}$$

Notice the similarity, and also that

- If $\theta = \theta_0$, then $\lambda = 0$ and Z_2^2 has a central chi-squared distribution.
- The probability of exceeding any critical value (power) can be made as large as desired by making λ bigger.
- There are 2 ways to make λ bigger.

Power calculation with R For $n = 100$, $\theta_0 = 0.50$ and $\theta = 0.60$

```
> # Power for Wald chisquare test of H0: theta=theta0
> n=100; theta0=0.50; theta=0.60
> lambda = n * (theta-theta0)^2 / (theta*(1-theta))
> critval = qchisq(0.95,1)
> power = 1-pchisq(critval,1,lambda); power
[1] 0.5324209
```

Earlier, had

```
> Z2power(0.60,100)
[1] 0.5324209
```

Check power calculations by simulation First develop and illustrate the code

```
# Try a simulation to test it.
set.seed(9999) # Set seed for "random" number generation
theta = 0.50; theta0 = 0.50; n = 100; m = 10
critval = qchisq(0.95,1); critval
p = rbinom(m,n,theta)/n; p
Z2 = sqrt(n)*(p-theta0)/sqrt(p*(1-p))
rbind(p,Z2)
sig = (Z2^2>critval); sig
sum(sig)/n
```

Output from the last slide

```
> # Try a simulation to test it.
> set.seed(9999) # Set seed for  "random" number generation
> theta = 0.50; theta0 = 0.50; n = 100; m = 10
> critval = qchisq(0.95,1); critval
[1] 3.841459
> p = rbinom(m,n,theta)/n; p
 [1] 0.40 0.56 0.47 0.57 0.47 0.50 0.58 0.48 0.40 0.53
> Z2 = sqrt(n)*(p-theta0)/sqrt(p*(1-p))
> rbind(p,Z2)
        [,1]     [,2]       [,3]     [,4]       [,5] [,6]     [,7]       [,8]      [,9]
p   0.400000 0.560000  0.4700000 0.570000  0.4700000  0.5 0.580000  0.4800000  0.400000
Z2 -2.041241 1.208734 -0.6010829 1.413925 -0.6010829  0.0 1.620882 -0.4003204 -2.041241
       [,10]
p  0.5300000
Z2 0.6010829
> sig = (Z2^2>critval); sig
 [1]  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE
> sum(sig)/n
[1] 0.02
```

Now the real simulation First estimated probability should equal about 0.05 because $\theta = \theta_0$

```
> # Check Type I error rate
> set.seed(9999)
> theta = 0.50; theta0 = 0.50; n = 100; m = 10000
> critval = qchisq(0.95,1)
> p = rbinom(m,n,theta)/n
> Z2 = sqrt(n)*(p-theta0)/sqrt(p*(1-p))
> sig = (Z2^2>critval)
> sum(sig)/m
[1] 0.0574

> # Power calculation for theta=0.60 said power = 0.5324209
> set.seed(9998)
> theta = 0.60; theta0 = 0.50; n = 100; m = 10000
> critval = qchisq(0.95,1)
> p = rbinom(m,n,theta)/n
> Z2 = sqrt(n)*(p-theta0)/sqrt(p*(1-p))
> sig = (Z2^2>critval)
> sum(sig)/m
[1] 0.5353
```

Conclusions from the power analysis

- Power with $n = 100$ is pathetic.
- As Fisher said, "To call in the statistician after the experiment is done may be no more than asking him to perform a postmortem examination: he may be able to say what the experiment died of."
- $n = 200$ is better.

  ```
  > n=200; theta0=0.50; theta=0.60
  > lambda = n * (theta-theta0)^2 / (theta*(1-theta))
  > power = 1-pchisq(qchisq(0.95,1),1,lambda); power
  [1] 0.8229822
  ```

- What sample size is required for power of 90%?

What sample size is required for power of 90%?

```
> # Find sample size needed for power = 0.90
> theta0=0.50; theta=0.60; critval = qchisq(0.95,1)
> effectsize = (theta-theta0)^2 / (theta*(1-theta))
> n = 0
```

```
> power=0
> while(power < 0.90)
+ {
+ n = n+1
+ lambda = n * effectsize
+ power = 1-pchisq(critval,1,lambda)
+ }
> n; power
[1] 253
[1] 0.9009232
```

General non-central chi-squared

Let $X_1, \ldots, X_n$ be independent $N(\mu_i, \sigma_i^2)$. Then

$$Y = \sum_{i=1}^{n} \frac{X_i^2}{\sigma_i^2} \sim \chi^2(n, \lambda), \text{ where } \lambda = \sum_{i=1}^{n} \frac{\mu_i^2}{\sigma_i^2}$$

- Density is a bit messy.
- Reduces to central chi-squared when $\lambda = 0$.
- Generalizes to $Y \sim \chi^2(\nu, \lambda)$, where $\nu > 0$ as well as $\lambda > 0$
- Stochastically increasing in λ, meaning $Pr\{Y > y\}$ can be increased by increasing λ.
- $\lim_{\lambda\to\infty} Pr\{Y > y\} = 1$
- Probabilities are easy to calculate numerically.