Economics 103 – Statistics for Economists

Francis J. DiTraglia

University of Pennsylvania

Lecture 20

An excerpt from The Lady Tasting Tea by David Salsburg

It was a summer afternoon in Cambridge, England, in the late 1920s. A group of university dons, their wives, and some guests were sitting around an outdoor table for afternoon tea. One of the women was insisting that tea tasted different depending upon whether the tea was poured into the milk or whether the milk was poured into the tea. The scientific minds among the men scoffed at this as sheer nonsense. What could be the difference? They could not conceive of any difference in the chemistry of the mixtures that could exist. A thin, short man, with thick glasses and a Vandyke beard beginning to turn gray, pounced on the problem. "Let us test the proposition" he said excitedly. He began to outline an experiment in which the lady who insisted there was a diference would be presented with a sequence of cups of tea, in some of which the milk had been poured into the tea and in others of which the tea had been poured into the milk.

Continued...

And so it was that summer afternoon in Cambridge. The man with the Vandyke beard was Ronald Aylmer Fisher, who was in his late thirties at the time. He would later be knighted Sir Ronald Fisher. In 1935, he wrote a book entitled The Design of Experiments, and he described the experiment of the lady tasting tea in the second chapter of that book. In his book, Fisher discusses the lady and her belief as a hypothetical problem. He considers the various ways in which an experiment might be designed to determine if she could tell the difference.

The Pepsi Challenge

(Volunteers: 1 "Expert," 1 Skeptic)

The Pepsi Challenge

Our expert claims to be able to tell the difference between Coke and Pepsi. Let's put this to the test!

- Eight cups of soda
 - ► Four contain Coke
 - Four contain Pepsi
- The cups are randomly arranged
- How can we use this experiment to tell if our expert can really tell the difference?

The Results:

of Cokes Correctly Identified:

What do you think? Can our expert really tell the difference?



- (a) Yes
- (b) No





• $\binom{8}{4} = 70$ ways to choose four of the eight cups.



- $\binom{8}{4} = 70$ ways to choose four of the eight cups.
- ▶ If guessing randomly, each of these is *equally likely*



- $\binom{8}{4} = 70$ ways to choose four of the eight cups.
- ▶ If guessing randomly, each of these is *equally likely*
- ▶ Only *one* of the 70 possibilities corresponds to correctly



- $\binom{8}{4} = 70$ ways to choose four of the eight cups.
- ▶ If guessing randomly, each of these is *equally likely*
- Only one of the 70 possibilities corresponds to correctly identifying all four cups of Coke.



- $\binom{8}{4} = 70$ ways to choose four of the eight cups.
- ▶ If guessing randomly, each of these is equally likely
- Only one of the 70 possibilities corresponds to correctly identifying all four cups of Coke.
- ▶ Thus, the probability is $1/70 \approx 0.014$





• $\binom{8}{4} = 70$ ways to choose four of the eight cups.



- $\binom{8}{4} = 70$ ways to choose four of the eight cups.
- ▶ If guessing randomly, each of these is *equally likely*



- $\binom{8}{4} = 70$ ways to choose four of the eight cups.
- ▶ If guessing randomly, each of these is *equally likely*
- ▶ There are 16 ways to mis-identify one Coke:
 - ▶ 4 choices of which Coke you call a Pepsi
 - ▶ 4 choices of which Pepsi you call a Coke
 - ▶ Total of $4 \times 4 = 16$ possibilities



- $\binom{8}{4} = 70$ ways to choose four of the eight cups.
- ▶ If guessing randomly, each of these is *equally likely*
- ▶ There are 16 ways to mis-identify one Coke:
 - ▶ 4 choices of which Coke you call a Pepsi
 - ▶ 4 choices of which Pepsi you call a Coke
 - ▶ Total of $4 \times 4 = 16$ possibilities
- ▶ Thus, the probability is $16/70 \approx 0.23$

Probabilities if Guessing Randomly

# Correct	0	1	2	3	4
Prob.	1/70	16/70	36/70	16/70	1/70



# Correct	0	1	2	3	4
Prob.	1/70	16/70	36/70	16/70	1/70

If you're just guessing, what is the probability of identifying *at least* three Cokes correctly?



# Correct	0	1	2	3	4
Prob.	1/70	16/70	36/70	16/70	1/70

If you're just guessing, what is the probability of identifying *at least* three Cokes correctly?

- Probabilities of mutually exclusive events sum.
- ▶ P(all four correct) = 1/70
- ▶ P(exactly 3 correct) = 16/70
- ▶ $P(\text{at least three correct}) = 17/70 \approx 0.24$

The Pepsi Challenge

- Even if you're just guessing randomly, the probability of correctly identifying three or more Cokes is around 24%
- In contrast, the probability of identifying all four Cokes correctly is only around 1.4% if you're guessing randomly.
- ▶ We should probably require the expert to get them all right.
- What if the expert gets them all wrong? This also has probability 1.4% if you're guessing randomly...

That was a Hypothesis Test!

We'll go through the details in a moment, but first an analogy...

Hypothesis Testing is Similar to a Criminal Trial

► The person on trial is either innocent or guilty (but not both!)

Hypothesis Testing

Either the null hypothesis H₀ or the alternative H₁ hypothesis is true.

- ► The person on trial is either innocent or guilty (but not both!)
- "Innocent Until Proven Guilty"

- ► Either the null hypothesis H_0 or the alternative H_1 hypothesis is true.
- ightharpoonup Assume H_0 to start

- The person on trial is either innocent or guilty (but not both!)
- "Innocent Until Proven Guilty"
- Only convict if evidence is "beyond a shadow of a doubt"

- Either the null hypothesis H₀ or the alternative H₁ hypothesis is true.
- ightharpoonup Assume H_0 to start
- Only reject H₀ in favor of H₁ if there is strong evidence.

- The person on trial is either innocent or guilty (but not both!)
- "Innocent Until Proven Guilty"
- Only convict if evidence is "beyond a shadow of a doubt"
- Not Guilty rather than Innocent

- Either the null hypothesis H₀ or the alternative H₁ hypothesis is true.
- ightharpoonup Assume H_0 to start
- Only reject H₀ in favor of H₁ if there is strong evidence.
- ► Fail to reject rather than Accept H₀

- The person on trial is either innocent or guilty (but not both!)
- "Innocent Until Proven Guilty"
- Only convict if evidence is "beyond a shadow of a doubt"
- Not Guilty rather than Innocent
 - ▶ Acquit ≠ Innocent

- Either the null hypothesis H₀ or the alternative H₁ hypothesis is true.
- ightharpoonup Assume H_0 to start
- Only reject H₀ in favor of H₁ if there is strong evidence.
- Fail to reject rather than Accept H₀
 - (Fail to reject H_0) \neq (H_0 True)

- The person on trial is either innocent or guilty (but not both!)
- "Innocent Until Proven Guilty"
- Only convict if evidence is "beyond a shadow of a doubt"
- Not Guilty rather than Innocent
 - ▶ Acquit ≠ Innocent
- Two Kinds of Errors:
 - Convict the innocent
 - Acquit the guilty

- Either the null hypothesis H₀ or the alternative H₁ hypothesis is true.
- ightharpoonup Assume H_0 to start
- Only reject H₀ in favor of H₁ if there is strong evidence.
- ► Fail to reject rather than Accept H₀
 - (Fail to reject H_0) \neq (H_0 True)
- Two Kinds of Errors:
 - ► Reject true *H*₀ (Type I)
 - ▶ Don't reject false H_0 (Type II)

- The person on trial is either innocent or guilty (but not both!)
- "Innocent Until Proven Guilty"
- Only convict if evidence is "beyond a shadow of a doubt"
- Not Guilty rather than Innocent
 - ► Acquit ≠ Innocent
- ► Two Kinds of Errors:
 - Convict the innocent
 - Acquit the guilty
- Convicting the innocent is a worse error. Want this to be rare even if it means acquitting the guilty.

- ► Either the null hypothesis H₀ or the alternative H₁ hypothesis is true.
- ightharpoonup Assume H_0 to start
- Only reject H₀ in favor of H₁ if there is strong evidence.
- ► Fail to reject rather than Accept H₀
 - (Fail to reject H_0) \neq (H_0 True)
- Two Kinds of Errors:
 - ► Reject true *H*₀ (Type I)
 - ▶ Don't reject false *H*₀ (Type II)
- Type I errors (reject true H₀) are worse: make them rare even if that means more Type II errors.

Null Hypothesis H_0

Can't tell the difference between Coke and Pepsi: just guessing.

Null Hypothesis H_0

Can't tell the difference between Coke and Pepsi: just guessing.

Alternative Hypothesis H_1

Able to distinguish Coke from Pepsi.

Null Hypothesis H_0

Can't tell the difference between Coke and Pepsi: just guessing.

Alternative Hypothesis H_1

Able to distinguish Coke from Pepsi.

Type I Error – Reject H_0 even though it's true

Decide expert can tell the difference when she's really just guessing.

Null Hypothesis H_0

Can't tell the difference between Coke and Pepsi: just guessing.

Alternative Hypothesis H_1

Able to distinguish Coke from Pepsi.

Type I Error – Reject H_0 even though it's true

Decide expert can tell the difference when she's really just guessing.

Type II Error – Fail to reject H_0 even though it's false

Decide expert just guessing when she really can tell the difference.

How do we find evidence to reject H_0 ?

- ▶ Choose a significance level α maximum probability of Type I error that we are willing to tolerate.
 - Measures how often we will reject a true null, i.e. convict an innocent person

- ▶ Choose a significance level α maximum probability of Type I error that we are willing to tolerate.
 - ► Measures how often we will reject a true null, i.e. convict an innocent person
- ▶ Test Statistic T_n uses sample to measure plausibility of H_0

- ▶ Choose a significance level α maximum probability of Type I error that we are willing to tolerate.
 - Measures how often we will reject a true null, i.e. convict an innocent person
- ▶ Test Statistic T_n uses sample to measure plausibility of H_0
- ▶ Null Hypothesis $H_0 \Rightarrow$ Sampling Distribution for T_n
 - "Under the null" means "assuming the H_0 is true"

- ▶ Choose a significance level α maximum probability of Type I error that we are willing to tolerate.
 - Measures how often we will reject a true null, i.e. convict an innocent person
- ▶ Test Statistic T_n uses sample to measure plausibility of H_0
- ▶ Null Hypothesis $H_0 \Rightarrow$ Sampling Distribution for T_n
 - "Under the null" means "assuming the H_0 is true"
- ▶ Using α and the sampling distribution of T_n under the null, we construct a decision rule in terms of a critical value c_{α}

- ▶ Choose a significance level α maximum probability of Type I error that we are willing to tolerate.
 - Measures how often we will reject a true null, i.e. convict an innocent person
- ▶ Test Statistic T_n uses sample to measure plausibility of H_0
- ▶ Null Hypothesis $H_0 \Rightarrow$ Sampling Distribution for T_n
 - ▶ "Under the null" means "assuming the H₀ is true"
- ▶ Using α and the sampling distribution of T_n under the null, we construct a decision rule in terms of a critical value c_{α}
 - Reject H_0 if $T_n > c_\alpha$

Test Statistic T_n

 $T_n =$ Number of Cokes correctly identified

 H_0 : No skill, just guessing randomly

Under this null hypothesis, the sampling distribution of T_n is:

# Correct	0	1	2	3	4
Prob.	1/70	16/70	36/70	16/70	1/70



 T_n : # of Cokes correctly identified. Sampling Dist. under H_0 :

# Correct	0	1	2	3	4
Prob.	1/70	16/70	36/70	16/70	1/70

If I choose a significance level of $\alpha=0.05$, what critical value should I use?

(Remember that α is the probability of rejecting H_0 when it is actually true.)



 T_n : # of Cokes correctly identified. Sampling Dist. under H_0 :

# Correct	0	1	2	3	4
Prob.	1/70	16/70	36/70	16/70	1/70

If I choose a significance level of $\alpha=0.05$, what critical value should I use?

(Remember that α is the probability of rejecting H_0 when it is actually true.)

Want $P(\text{Reject } H_0|H_0 \text{ True}) \leq 0.05$ $P(T_n \geq 3|\text{Just Guessing}) = 17/70 \approx 0.23 > 0.05$ $P(T_n \geq 4|\text{Just Guessing}) = 1/70 \approx 0.014 \leq 0.05$



 T_n : # of Cokes correctly identified. Sampling Dist. under H_0 :

# Correct	0	1	2	3	4
Prob.	1/70	16/70	36/70	16/70	1/70

If I choose a significance level of $\alpha=0.25$, what critical value should I use?



 T_n : # of Cokes correctly identified. Sampling Dist. under H_0 :

# Correct	0	1	2	3	4
Prob.	1/70	16/70	36/70	16/70	1/70

If I choose a significance level of $\alpha=0.25$, what critical value should I use?

Want
$$P(\text{Reject } H_0|H_0 \text{ True}) \leq 0.25$$

 $P(T_n \geq 2|\text{Just Guessing}) = 53/70 \approx 0.76 > 0.25$
 $P(T_n \geq 3|\text{Just Guessing}) = 17/70 \approx 0.23 \leq 0.25$



 H_0 : Expert is just guessing randomly.

 H_1 : Expert can distinguish Coke from Pepsi.

 T_n : # of Cokes correctly identified. Has following sampling under the null:

# Correct	0	1	2	3	4
Prob.	1/70	16/70	36/70	16/70	1/70

If I choose $\alpha = 0.05$, what decision rule should I use?

- (a) Reject H_0 if $T_n \ge 0$
- (b) Reject H_0 if $T_n \geq 1$
- (c) Reject H_0 if $T_n \geq 2$
- (d) Reject H_0 if $T_n \geq 3$
- (e) Reject H_0 if $T_n \ge 4$

 H_0 : Expert is just guessing randomly.

 H_1 : Expert can distinguish Coke from Pepsi.

 T_n : # of Cokes correctly identified. Has following sampling under the null:

# Correct	0	1	2	3	4
Prob.	1/70	16/70	36/70	16/70	1/70

If I choose $\alpha = 0.05$, what decision rule should I use?

Need
$$P(\text{Reject } H_0|H_0 \text{ True}) \leq \alpha = 0.05$$

$$P(T_n \ge 3 | \text{Just Guessing}) = 17/70 \approx 0.23 > 0.05$$

$$P(T_n \ge 4|\text{Just Guessing}) = 1/70 \approx 0.014 \le 0.05$$

Critical value for $\alpha = 0.05$ is 4



 H_0 : Expert is just guessing randomly.

 H_1 : Expert can distinguish Coke from Pepsi.

 T_n : # of Cokes correctly identified. Has following sampling under the null:

# Correct	0	1	2	3	4
Prob.	1/70	16/70	36/70	16/70	1/70

If I choose $\alpha = 0.25$, what critical value should I use?

- (a) 0
- (b) 1
- (c) 2
- (d) 3
- (e) 4