Day 7 Variations on Test T+ (new ex starts p. 29)

1-sided normal testing

H_0: μ ≤ 150 vs. H_1: μ > 150 (Let σ = 10, n = 100)

let significance level α = .025

Reject H_0 whenever M ≥ 150 + 2σ/√n: M ≥ 152

M is the sample mean, its value is M_0.
1SE = σ/√n = *1*

Rejection rules:

Reject iff M > 150 + 2SE (N-P)

In terms of the P-value:

Reject iff P-value ≤ .025 (Fisher)

(P-value a distance measure, but inverted)

Let M = 152, so I reject H_0.

PRACTICE WITH P-VALUES
Let M = 152

$Z = (152 – 150)/1 = 2$

The P-value is $Pr(Z > 2) = .025$

PRACTICE WITH P-VALUES

Let M = 151

$Z = (151 – 150)/1 = 1$

The P-value is $Pr(Z > 1) = .16$

$SEV(\mu > 150) = .84 = 1 – \text{P-value}$

PRACTICE WITH P-VALUES

Let M = 150.5

Z = (150.5 – 150)/1 = .5

The P-value is Pr(Z > .5) = .3

PRACTICE WITH P-VALUES

Let M = 150

$Z = (150 – 150)/1 = 0$

The P-value is $Pr(Z > 0) = .5$

Frequentist Evidential Principle: FEV

FEV (i). ***x*** is evidence against H_0 (i.e., evidence of discrepancy from H_0), if and only if the P-value $\Pr(d > d_0; H_0)$ is very low (equivalently, $\Pr(d < d_0; H_0) = 1 - P$ is very high).

Contraposing FEV(i) we get our minimal priniciple

FEV (ia) ***x*** are poor evidence against H_0 (poor evidence of discrepancy from H_0), if there's a high probability the test would yield a more discordant result, if H_0 is correct.

Note the one-directional 'if' claim in FEV (1a)

(i) is not the only way **x** can be BENT.

Reformulating Tests: P-values Don't Give an Effect Size

Severity function: SEV(Test T, data ***x***, claim *C*)

- Tests are reformulated in terms of a discrepancy γ from H_0
- Instead of a binary cut-off (significant or not) the particular outcome is used to infer discrepancies that are or are not warranted

H_0: μ ≤ 150 vs. H_1: μ > 150 (Let σ = 10, *n* = 100)

The usual test infers there's an indication of *some* positive discrepancy from 150 because

$$Pr(M < 152: H_0) = .97$$

Not very informative

Are we warranted in inferring μ > 153 say?

- Recall the complaint of the Likelihoodist (p. 36)

- For them, inferring H_1: $\mu > 150$ means every value in the alternative is more likely than 150

- Our inferences are not to point values, but we agree to the need to block inferences to discrepancies beyond those warranted with severity.

Consider: How severely has $\mu > 153$ passed the test?

SEV(μ > 153) (p. 143)

M = 152, as before, claim *C*: $\mu > 153$

The data "accord with C" but there needs to be a reasonable probability of a worse fit with C, if C is false

Pr(*"a worse fit"*; *C* is false)

Pr(M ≤ 152; μ ≤ 153)

Evaluate at $\mu = 153$, as the prob is greater for $\mu < 153$.

Consider: How severely has $\mu > 153$ passed the test?

To get $Pr(M \leq 152: \mu = 153)$, standardize:

$Z = \sqrt{100}\,(152 - 153)/1 = -1$

$Pr(Z < -1) = .16$ Terrible evidence

Consider: How severely has $\mu > 150$ passed the test?

To get $\Pr(M \leq 152: \mu = 150)$, standardize:
$Z = \sqrt{100}\,(152 - 150)/1 = 2$

$\Pr(Z < 2) = .97$

Notice it's 1 – P-value

Now consider SEV(μ > 150.5) (still with M = *152)*

Pr (A worse fit with *C*; claim is false) = .97

Pr(M < 152; μ = 150.5)

Z = (152 – 150.5) /1 = 1.5

Pr (Z < 1.5)= .93 Fairly good indication μ > 150.5

Table 3.1 Reject in test T+: H_0: $\mu \leq 150$ vs. H_1: $\mu > 150$ with $\bar{x} = 152$

Claim $\mu > \mu_1$	Severity $\Pr(\bar{X} \leq 152; \mu = \mu_1)$
$\mu > 149$	0.999
$\mu > 150$	0.97
$\mu > 151$	0.84
$\mu > 152$	0.5
$\mu > 153$	0.16

$\mu > 150.5$ → .093

FOR PRACTICE:

Now consider SEV(*μ* > 151)   (still with *M = 152)*

Pr (A worse fit with *C*; claim is false) = __

Pr(M < 152; μ = 151)

Z = (152 – 151) /1 = 1

Pr (Z < 1)= .84

MORE PRACTICE:

Now consider SEV(μ > 152)   (still with M = 152)

Pr (A worse fit with *C*; claim is false) = __

Pr(M < 152; μ = 152)

Z = 0

Pr (Z < 0)= .5–important benchmark

Terrible evidence that μ > 152

Table 3.2 has exs with M = 153.

Compare n = 100 with n = 10,000

H_0: $\mu \leq 150$ vs. H_1: $\mu > 150$ (Let $\sigma = 10$, n = 10,000)

Reject H_0 whenever M ≥ 2SE: M ≥ 150.2

M is the sample mean (significance level = .025)

1SE = $\sigma/\sqrt{n} = 10/\sqrt{10{,}000} = .1$

Let M = 150.2, so I reject H_0.

Comparing n = 100 with n = 10,000

Reject H_0 whenever M ≥ 2SE: M ≥ 150.2

$SEV_{10,000}$(μ > 150.5) = 0.001

Z = (150.2 – 150.5) /.1 = -.3/.1 = -3
P(Z < -3) = .001

Corresponding 95% CI: [0, 150.4]

A .025 result is terrible indication μ > 150.5
When reached with n = 10,000

While SEV_{100}(μ > 150.5) = 0.93

***Non-rejection*.** Let M = 151, the test does not reject H_0.

The standard formulation of N-P (as well as Fisherian) tests stops there.

We want to be alert to a fallacious interpretation of a "negative" result: inferring there's no positive discrepancy from μ = 150.

Is there evidence of compliance? μ ≤ 150?
The data "accord with" H_0, but what if the test had little capacity to have alerted us to discrepancies from 150?

No evidence against H_0 is not evidence for it.

Condition (S-2) requires us to consider $\Pr(X > 151; 150)$, which is only .16.

P-value "moderate"

FEV(ii): A moderate *p* value is evidence of the absence of a discrepancy γ from H_0, only if there is a high probability the test would have given a worse fit with H_0 (i.e., smaller *P*- value) were a discrepancy γ to exist.

For a Fisherian like Cox, a test's power only has relevance pre-data, they can measure "sensitivity".

> In the Neyman-Pearson theory of tests, the sensitivity of a test is assessed by the notion of *power*, defined as the probability of reaching a preset level of significance …for various alternative hypotheses. In the approach adopted here the assessment is via the distribution of the random variable *P*, again considered for various alternatives (Cox 2006, p. 25)

Computation for SEV(T, M = 151, *C*: μ ≤ 150)
Z = (151 – 150)/1 = 1

Pr(Z > 1) = .16

SEV(*C*: μ ≤ 150) = low (.16).

- So there's poor indication of H_0

Refers to Table 3.3 p. 145

Can they say M = 151 is a good indication that $\mu \leq 150.5$?

No, SEV(T, M = 151, C: $\mu \leq 150.5$) = ~.3.

[Z = 151 – 150.5 = .5]

But M = 151 is a good indication that $\mu \leq 152$

[Z = 151 – 152 = -1; Pr (Z > -1) = .84]

SEV($\mu \leq 152$) = .84

It's an even better indication $\mu \leq 153$ (Table 3.3, p. 145)

[Z = 151 – 153 = -2; Pr (Z > -2) = .97]

$\Pi(\gamma)$: "sensitivity function"

Computing $\Pi(\gamma)$ views the P-value as a statistic.
$\Pi(\gamma) = \Pr(P < p_{obs}; \mu_0 + \gamma)$.

The alternative $\mu_1 = \mu_0 + \gamma$.

Given that P-value inverts the distance, it is less confusing to write $\Pi(\gamma)$

$\boldsymbol{\Pi(\gamma) = \Pr(d > d_0; \mu_0 + \gamma)}$.

Compare to the power of a test:

$POW(\gamma) = \Pr(d > c_\alpha; \mu_0 + \gamma)$ the N-P cut-off c_α.

FEV(ii) in terms of $\Pi(\gamma)$

P-value is modest *(not small):* Since the data accord with the null hypothesis, FEV directs us to examine the probability of observing a *result more discordant from H_0* if $\mu = \mu_0 + \gamma$:

If $\Pi(\gamma) = \Pr(d > d_0; \mu_0 + \gamma)$ is very high, the data indicate that $\mu < \mu_0 + \gamma$.

Here $\Pi(\gamma)$ gives the severity with which the test has probed the discrepancy γ.

FEV (ia) in terms of *Π*(γ)

If *Π*(γ) = Pr(d > do; μ_0 +γ) = moderately high (greater than .3, .4, .5), then there's poor grounds for inferring $\mu > \mu_0 + \gamma$.

This is equivalent to saying the SEV($\mu > \mu_0 + \gamma$) is poor.

New Example for Day 7: August 3

H_0: $\mu \leq 150$ vs. H_1: $\mu > 150$ (Let $\sigma = 10$, $\boldsymbol{n} = \mathbf{25}$)

let significance level $\alpha = .025$

Reject iff M > 150 + 2SE (N-P)

Reject H_0 whenever $M \geq 150 + 2\sigma/\sqrt{n}$: $M \geq 152$

M is the sample mean, its value is M_0.

$1SE = \sigma/\sqrt{25} = 10/5 = 2$

Reject H_0 whenever $M \geq 150 + 2(2)$: $M \geq 154$

Let $M_0 = 154$, just at the .025 cut-off.

Assess SEV for the same claims in Table 3.1 p. 144

Claim $\mu > 149$

SEV $(\mu > 149) = \Pr(M \leq 154; 149)$

$Z = (154 – 149)/2 = 2.5$

$\Pr(Z \leq 2.5) = .99$

Now consider

Claim $\mu > 150$ (for the same outcome $M = 154$)

$SEV(\mu > 150) = Pr(M \leq 154; 150)$

$Z = (154 - 150)/2 = 2$

$Pr(Z \leq 2) = .97$

Now consider

Claim $\mu > 151$ (for the same outcome $M = 154$)

SEV ($\mu > 151$) = $Pr(M \leq 154; 151)$

$Z = (154 - 151)/2 = 1.5$

$Pr(Z \leq 1.5) = .93$

Now consider

Claim $\mu > 152$ (for the same outcome $M = 154$)

$SEV(\mu > 152) = Pr(M \leq 154; 152)$

$Z = (154 - 152)/2 = 1$

$Pr(Z \leq 1) = .84$

Now consider

Claim $\mu > 153$ (for the same outcome $M = 154$)

$SEV(\mu > 153) = Pr(M \leq 154; 153)$

$Z = (154 - 153)/2 = .5$

$Pr(Z \leq .5) = .69$

Add one beyond that table

Claim $\mu > 154$ (for the same outcome $M = 154$)

$SEV(\mu > 154) = Pr(M \leq 154; 154)$

$Z = (154 - 154)/2 = 0$

$Pr(Z \leq 0) = .5$

The warrant gets worse and worse for larger discrepancies given the same outcome

Use this example for Power (or prob of Type II errors)

H_0: μ ≤ 150 vs. H_1: μ > 150 (Let σ = 10, ***n* = 25**)

let significance level α = .025

Since we let M be just at the cut-off for rejection, we can use the same computations to get the power of the test against different values

(SIST focuses in on power in Excursion 5, but since we've introduced it, might as well use the time to look at it)

Reject H_0 whenever $M \geq 150 + 2(2)$: $M \geq 154$

Power of test T+ against $\mu' = POW(\mu')=$

Pr (Test T+ would reject μ_0 when were $\mu = \mu'$)

$Pr(M \geq 154; \mu')$

$Z = (154 - \mu')/2$

We're just changing different hypothesized values for μ'.

Power of test T+ against $\mu = 150$ = POW(150)=

Pr($M \geq 154$; 150)

$Z = (154 – 150)/2 = 2$ (1.96 is the official cut-off)

Pr($Z \geq 2$) = .025 (I often round to .030

Since it’s continuous can use > or ≥

Note it’s = alpha

POW(151)=

Pr(M ≥ 154; 151)

Z = (154 – 151)/2 = 1.5

Pr(Z ≥ 1.5) = .07

POW(152)=

Pr(M ≥ 154; 152)

Z = (154 – 152)/2 = 1

Pr(Z ≥ 1.5) = .16

You see these values are the complements of the corresponding SEV for claiming to have evidence that μ > μ’
For each value of μ’

POW(153)=

Pr(M ≥ 154; 153)

Z = (154 – 153)/2 = .5

Pr(Z ≥ .5) = .31

You see the power to detect discrepancies is increasing, the larger the alternative. It’s still not even .5.

POW(154)=

Pr(M ≥ 154; 154)

Z = (154 – 154)/2 = 0

Pr(Z ≥ 0) = .5

Finally, the alternative just equal to the cut-off is .5.

Jump to

POW(156)=

156 = 1 SE greater than the null value of 150

Pr(M ≥ 154; 156)

Z = (154 – 156)/2 = -1

Pr(Z ≥ -1) = .84

An important benchmark for test T+.

Spoze someone required $\alpha = .02$

$H_0: \mu \leq 150$ vs. $H_1: \mu > 150$ (Let $\sigma = 10$, ***n* = 25**)

let significance level $\alpha = .02$

Then you need to reach 2.1SE to find stat significance, M would need to be 154.2

So observing M = 154, you have a “non-reject” if you use cut-offs.

We’re just using this example for practice with computations, taking advantage of numbers we already have.

Reject H_0 whenever $M \geq 150 + 2.1(2)$: $M \geq 154.2$

Not very informative to just say non-reject, and fallacious to say evidence of no difference from 150

We can set upper bounds that are warranted with reasonable severity and those that are not

Note the SEV values don't change because it considers M, not the cut-off)

SEV(Test T+, M, $(\mu > 150 + \gamma)$)

$SEV(\mu > 150) = .97$

$SEV(\mu > 151) = .93$

$SEV(\mu > 152) = .84$

$SEV(\mu > 153) = .69$

So, since the two claims form a partition of the parameter space

SEV($\mu \leq 150$) = .03

SEV($\mu \leq 151$) = .07

SEV($\mu \leq 152$) = .16

SEV($\mu \leq 153$) = .31

____

SEV($\mu \leq 154$) = .5

SEV($\mu \leq 155$) = .69

SEV($\mu \leq 156$) = .84 our benchmark for decent SEV

POW(152)=

Pr(M ≥ 154; 152)

Z = (154 – 152)/2 = 1

Pr(Z ≥ 1.5) = .16

You see these values are the complements of the corresponding SEV for claiming to have evidence that μ > μ’
For each value of μ’

POW(153)=

Pr(M ≥ 154; 153)

$Z = (154 - 153)/2 = .5$

$Pr(Z \geq .5) = .31$

You see the power to detect discrepancies is increasing, the larger the alternative. It's still not even .5.

POW(154)=

Pr(M ≥ 154; 154)

Z = (154 – 154)/2 = 0

Pr(Z ≥ 0) = .5

Finally, the alternative just equal to the cut-off is .5.

Jump to

POW(156)=

156 = 1 SE greater than the null value of 150

Pr(M ≥ 154; 156)

Z = (154 – 156)/2 = -1

Pr(Z ≥ -1) = .84

An important benchmark for test T+.