11/12/2019 ①

Summary of Hypothesis Testing and Schuams Statistics Ch 10

Ingredients for Hypoth testing

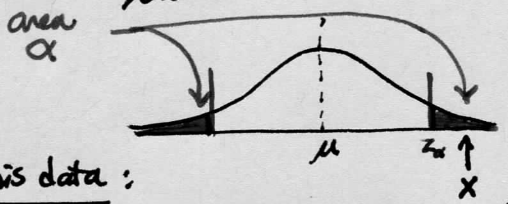
Fisher

RV: we have an experiment which generates a random variable X

Distribution: X is associated with a 1-param family of distribs, say $X \sim Bin(n, p)$ where n would be fixed and p parameterizes the family.
[Schuams always approximates $Bin(n,p) \sim \mathcal{N}(\mu, \sigma^2)$ where $\mu = np$, $\sigma = \sqrt{np(1-p)}$]

Null Hypoth H_0: a specific claim like $p = 1/2$ (which determines a particular distrib $X \sim Bin(n, 1/2)$ say (n fixed)). We attempt to reject H_0 if the "data makes it unlikely" – more on this concept later in the section on objections to hypoth testing. Note also that we don't accept H_0, only reject or fail to reject it. Presumably because another hypoth could also produce the same data.

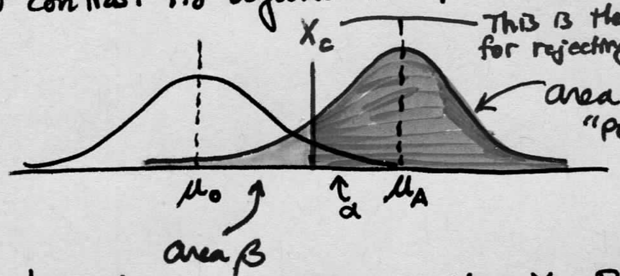
Significance Level α: a pre-chosen bound that defines what we mean that the data is unlikely for H_0. If the RV X lies in the tails of area α, then we reject H_0
[If H_0 would be true, then α = probability of false reject]



p-Value of this data:

Given X and H_0, this is the smallest value of α that would cause us to reject H_0.

Neyman-Pearson Now contrast H_0 against a specific alternative H_A (another member of 1-param family NOT everything like $-H_0$)



"Imperfect Decision Criteria" is my term

Given an alternative hypoth H_A: $p = p_A$ and thus $X \sim Bin(n, p_A)$ (n fixed) we have the H_A pdf and the cutoff value (X_c) for α also cuts the H_A pdf. If H_A would be true and we do this test for H_0 with this cutoff value, β = probability we falsely "accept" H_0 and Power $= (1-\beta)$ is prob we correctly reject H_0 in favour of H_A.

- Note that as $\mu_A \to \mu_0$ the power decreases to min value α when bumps coincide.
- Note that for fixed μ_0 and μ_A, if we move the cutoff line to decrease α, we increase β and vice versa.

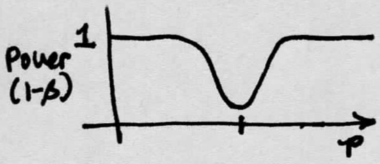
	Ho true	H_A true
IDC says do not reject Ho from this data		Type II error False accept H_0 Prob $= \beta$
IDC says reject H_0	Type I error False reject Prob $= \alpha$	"Power" of IDC Prob $= 1-\beta$

If H_A is true, the probability of detecting H_A is true with this IDC

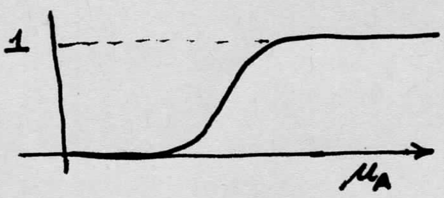
11/12/2019
(2)

There is more to Neyman-Pearson than a single H_A. The pdf for H_A is parameterized as a family of distributions, so we can plot the power of the IDC wrt this parameter.

In prob (12) we see



and in problem (14)



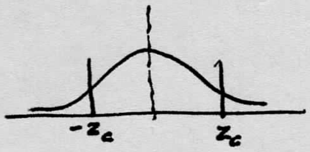
This spectrum of values gives the best idea of how the test behaves and where it is effective.

Problem (15)
By increasing the sample size n, we can attain arb small α and β

▷ Hypoth Testing and Confidence Intervals prob (16)

To form CI, we have an unknown pop μ and we compute $\bar{x}$ from sample of size n

we assume $f_{\bar{x}} \approx \mathcal{N}(\mu, \sigma^2/n)$. Form $Z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$.



we have $-z_c < \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}} < z_c \xRightarrow{\text{unnormalize}} \bar{x} - z_c\dfrac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_c\dfrac{\sigma}{\sqrt{n}}$

$\Rightarrow \mu \in \left(\bar{x} - z_c\dfrac{\sigma}{\sqrt{n}}, \bar{x} + z_c\dfrac{\sigma}{\sqrt{n}}\right)$

For Hypoth Testing (for the mean) again we use $Z = \dfrac{\bar{x}-\mu}{\sigma_{\bar{x}}} = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

We "accept" H_0 if $z \in (-z_c, z_c)$ for a 2-tailed test,

so again unnormalize $-z_c < \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}} < z_c$

But here we solve for $\bar{x}$ rather than μ: $\bar{x} \in \left(\mu - z_c\dfrac{\sigma}{\sqrt{n}}, \mu + z_c\dfrac{\sigma}{\sqrt{n}}\right)$

NOT nec the same interval.

▷ Criticism of Hypoth Testing from Jacob Cohen 'The Earth is Round ($p < 0.05$)'

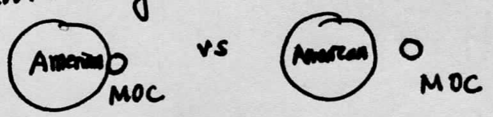
- Rules of logic don't nec hold for 'probably'

$A \Rightarrow B$ but we can't say $A \xRightarrow{\text{Prob}} B$
$-B \Leftarrow -B$ $-A \Leftarrow -B$

(I am American) $\xRightarrow{\text{Prob}}$ (I am NOT a Member of Congress)

(I am NOT American) $\not\Leftarrow$ (I am a MOC) FALSE

we can't distinguish



- Let $H = H_0$ is true D = Experiment gave this data

we test $P(D|H)$ but we want $P(H|D)$

"Given this ~~data~~ H_0, is there a low prob of this data?" NOT THE SAME "Given this data, is there a low prob of H_0?"

Elementary Bayes:
$$P(A|B) = \frac{P(A)\,P(B|A)}{P(B)}$$

$P(D|H) \neq P(H|D)$

It seems to me though, that it must be possible to show the logic of HT is valid at least in elementary cases of say binomial thm.

Want: $P(D|H)$ small $\Rightarrow P(H|D)$ small

How to evaluate $P(H)$? How could we say $P(H) = P(D)$?

- "The 'Nil' hypoth is always false" Cohen claims that if H_0 is something like $A = B$, if we run the sample size n high enough, some small difference will always appear

Schuam's statistics ch 10

p.171

NOT UNTIL PROB 3 DO WE REALLY GET STARTED

11/2/2017

① Flip a coin 100 times. Let $X = \#$ of heads. $n=100$ $p=\frac{1}{2}$ $X \sim Bin(n,p)$

Here we know this coin is fair.

Find $P(40 \le X \le 60)$

we know $P(X=40) = \binom{100}{40} p^{40}(1-p)^{60}$

$P(X=41) = \binom{100}{41} p^{41}(1-p)^{59}$

$$\Rightarrow P(40 \le X \le 60) = \sum_{i=40}^{60} P(X=i) = \sum_{i=40}^{60} \binom{100}{i} \underbrace{p^i (1-p)^{100-i}}_{(\frac{1}{2})^i (\frac{1}{2})^{100-i} = (\frac{1}{2})^{100}}$$

$$= \frac{1}{2^{100}} \sum_{i=40}^{60} \binom{100}{i}$$

$$\left.\begin{aligned} np &= 100\cdot\tfrac{1}{2} = 50 > 5 \\ n(1-p) &= 100\cdot\tfrac{1}{2} = 50 > 5 \end{aligned}\right\}$$ by p.124 we can use Normal approx.

$\mu := np = 50$

$\sigma := \sqrt{np(1-p)} = \sqrt{100\cdot\frac{1}{2}\cdot\frac{1}{2}} = \sqrt{25} = 5$

Convert to continuum: $39.5 \le X \le 60.5$

$$\frac{39.5-50}{5} \le \frac{X-\mu}{\sigma} \le \frac{60.5-50}{5}$$

$$-2.1 \le Z \le 2.1$$

Do lookup in Normal table (p.343) $\Rightarrow P(-2.1 \le Z \le 2.1) = 0.9642$

□

② Now flip a new coin 100 times. Is it fair or biased?

Hypoth H_0 = coin is fair $(p=\frac{1}{2})$ "Imperfect Decision criteria"

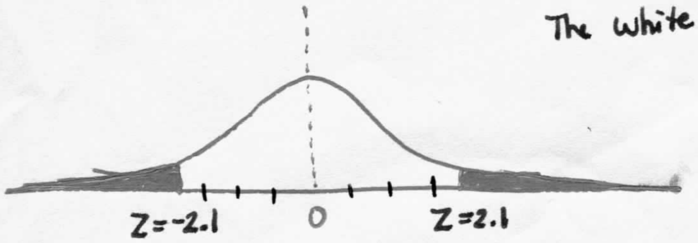
We decide we don't reject H_0 if $X \in [40,60]$

ⓐ Find prob of H_0 being rejected when it is really true. Type I Error "False Reject H_0"

$$P(X>60) + P(X<40) = 1 - P(40 \le X \le 60)$$
$$= 1 - 0.9642 \text{ from above}$$
$$= 0.0358$$

we reject at 3.58% level of significance.

ⓑ we can show this graphically:



The white area corresponds to $-2.1 \le Z \le 2.1$

i.e. $\underline{40 \le X \le 60}$ (well, $39.5 \le X \le 60.5$)

Thus, given say $X=53$ we can compute $Z = \frac{53-50}{5} = 0.6$ and see this lies in the white region, where we do not reject H_0.

(Converting to Z values is more useful in other problems)

Schuams ch 10 10/19/2019

③ Flip a coin $N=64$ times and test the hypoth that the coin is fair.
[Here we just set up test criteria]

Ingredients
1. RV: $X = \#$ of heads
2. Distrib: $X \sim Bin(64, p)$ which we will approx by $\mathcal{N}$ since N large enough
3. H_0 : $p = 1/2$
4. Level of signif: Set $\alpha = 0.05$. This is the area of the tails of distrib outside our cutoff lines. We reject H_0 if X lies out here. If H_0 would be true, the Prob that we would get a value in this region is α. False reject

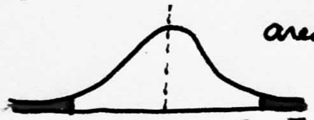
Get set up for $\mathcal{N}$ approx: (assuming H_0)

$\mu = Np = 64 \cdot \frac{1}{2} = 32$

$\sigma = \sqrt{Np(1-p)} = \sqrt{64 \cdot \frac{1}{2} \cdot \frac{1}{2}} = 4$

$$Z = \frac{X-\mu}{\sigma} = \frac{X-32}{4}$$

Do $\mathcal{N}$ approx



area $\alpha = 0.05 \xrightarrow{\text{2 tail test}} \left(\hat{\alpha} = \frac{1}{2}\alpha = 0.025\right) \xrightarrow{\text{table only gives right side}} \frac{1}{2} - \hat{\alpha} = 0.4750 \xrightarrow{\text{table lookup}} z_c = 1.96$

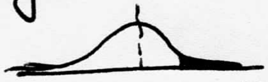
$\Rightarrow -1.96 \le \frac{X-32}{4} \le 1.96$

$\Rightarrow 24.16 \le X \le 39.85$ So we reject H_0 if $X > 39.85$ or $X < 24.16$. This is what I call the 'Imperfect Decision Criteria' or 'IDC'

---

⑤ Testing for ESP. The subject guesses the color of $N=50$ red/black cards. She guesses 32 correctly? Does this imply ESP?

Ingredients
- $X := \#$ of correct guesses, here 32
- Family of distribs: $X \sim Bin(50, p)$
- H_0: $p = 1/2$ [NO ESP] Thus H_0 specifies $X \sim Bin(50, 1/2)$ a particular member of family. We are interested in contrasting this with only $p > 1/2$ so we choose a 1-tail test.



- Based on this data ($X=32$) can we reject H_0 at $\alpha = 0.05$? $\alpha = 0.01$?

Differing from Schuam's, I will compute the p-value with makes this easier.

Set up for $\mathcal{N}$ approx:

$\mu = Np = 50 \cdot \frac{1}{2} = 25$

$\sigma = \sqrt{Np(1-p)} = 3.54$

$$Z = \frac{X-\mu}{\sigma} \qquad z_0 = \frac{32-25}{3.54} = 1.98$$

Compute p-Value

$\alpha = \hat{\alpha} \xleftarrow{\text{1 tail}} \frac{1}{2} - \hat{\alpha} = 0.4761 \longleftarrow .4761 \xleftarrow{\text{reverse table lookup}} z = 1.98$

$\Rightarrow \hat{\alpha} = 0.0239$

$\alpha = 0.0239$

$\boxed{\text{p-Value} = 0.0239}$ for this data $X=32$

Thus we can reject at level $\alpha = 0.01$ but not at $\alpha = 0.05$: $0.01 < \text{p-Value} < 0.05$

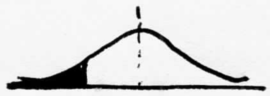
Rice MSADA p.283 defines p-value as the smallest level of signif for which H_0 can be rejected. □

Schuams
ch 10

10/19/2019

(6) Medicine claims to be "90% effective". We do a trial of $N=200$ people and it works for 160 of them.

Ingredients
- RV: X = # of people who had success, $X=160$
- Distrib family: $X \sim Bin(200, p)$
- H_0: $p=0.9$ $\Rightarrow X \sim Bin(200, 0.9)$. we are interested in $p<0.9$ so we choose a 1-tail test
- we set signif level $\alpha=0.01$



Set up $\mathcal{N}$ approx

$\mu = Np = 200 \cdot \frac{9}{10} = 180$

$\sigma = 4.23$

$[160 \rightsquigarrow (159.5, 161.5)]$

we don't bother with "continuity correction" on X, just take $Z_0 = \dfrac{160-180}{4.23} = -4.73$

compute p-Value

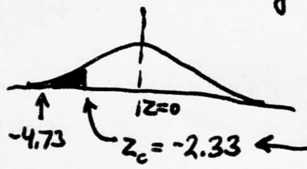
we can try to reverse lookup $^{+}4.73$ in table p.343: But every value >3.9 maps to 0.500

$\alpha = \frac{1}{2} - 0.500 \leftsquigarrow 0.500 \leftsquigarrow +4.73$ (symm)

$\alpha = 0$

Thus we get a p-Value of 0 which is certainly less than $\alpha=0.01$ and we reject H_0

▷ Since I don't really like probability 0 due to rounding error, we can also do it the book's way:



This is the cutoff for $\alpha=0.01$ and we see -4.73 lies in the tail and we reject H_0

□

(7) For a sample of $n=100$ light bulbs, we have mean lifetime 1570 hrs with std dev 120. Test hypoth for population $\mu=1600$ against $\mu \neq 1600$ for level of signif $\alpha=0.05$ and $\alpha=0.01$.

Ingredients X = lifetime of bulb $X \sim \mathcal{N}(\mu, \sigma^2)$ we don't know this?

Our RV: $\bar{X}$ = mean lifetime of sample $n=100$

Distrib: $\bar{X} \sim f_{\bar{X}}$ which we can approx by $\mathcal{N}(\mu_{\bar{X}}, \sigma_{\bar{X}})$ by ch 8,9 work [so again we get to $\mathcal{N}$ approx as we always do]

H_0: $\mu=1600$ against $\mu \neq 1600$ so use 2-tail test

set up $\mathcal{N}$ approx

$$Z_0 = \frac{\bar{X}-E(\bar{X})}{\sqrt{Var(\bar{X})}} = \frac{\bar{X}-\mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{1570-1600}{120/\sqrt{100}} = \frac{-30}{12} = -2.5$$

Should use t-distrib? No, Sample size $n \geq 30$

Compute p-Value

$$Z_0 = -2.5 \xrightarrow[\text{reverse table lookup}]{} 0.4938 \xrightarrow[\text{tail area}]{} \tfrac{1}{2} - 0.4938 = \underbrace{0.0062}_{\hat{\alpha}} \xrightarrow[\alpha = 2\hat{\alpha}]{\text{2 tail}} \text{p-Value} = 0.0124$$

$0.01 < 0.0124 < 0.05$ so we can reject H_0 at $\alpha=0.05$ but not at $\alpha=0.01$

□

10/19/2019

Now we introduce Neyman-Pearson paradigm and compare H_0 against a specific H_A

Flipping a coin again, do prob of heads $\underbrace{p=1/2}_{H_0}$ or $\underbrace{p=7/10}_{H_A}$? Specifically they are asking us to find the prob of failing to reject H_0 when H_A is true. [This prob is called β].

Define ingredients

- Number of trials $N=100$
- RV X = # of heads
- Distrib family $X \sim \mathrm{Bin}(100, p)$
- $H_0: p = 1/2 \Rightarrow X \sim \mathrm{Bin}(100, 1/2) \Rightarrow \begin{matrix}\mu = 50 \\ \sigma = 5\end{matrix}$

This is 'IDC'

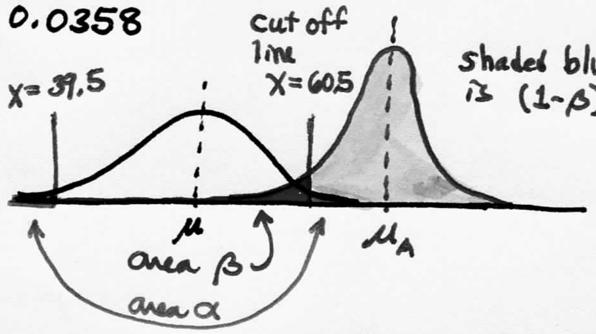
Alternative — another member of family

$H_A: p = 0.7 \Rightarrow X \sim \mathrm{Bin}(100, .7)$

$\mu_A = 100(.7) = 70$

$\sigma_A = \sqrt{100 \cdot (.7)(.3)} = 4.58$

The book declares we will reject H_0 if $X < 39.5$ or $X > 60.5$. They got these values from problems ① and ②, but it is cleaner if we just realize they come from a 2-tailed test with $\alpha = 0.0358$



The area β is the prob of failing to reject H_0 when H_A is actually true.

$\boxed{\text{Power} := (1-\beta)}$ Prob of data implying H_A when H_A is true.

▷ You can see that as $\mu_A \to \mu$ the power decreases (blue area decreases). When $\mu_A = \mu$ then we only would choose H_A if data point in the α tails of $H_0 \Rightarrow 1-\beta = \alpha$ here. See prob ⑫

▷ For a fixed μ_A, if we moved the cutoff line to decrease α, we decrease $(1-\beta)$ too i.e. we increase error β.

▷ Lets calculate β. Just like every other problem here, we must map to $N(0,1)$, but now we map the pdf for H_A: $39.5 < X < 60.5$

$$Z = \frac{X-\mu_A}{\sigma_A} \Rightarrow \frac{39.5-\mu_A}{\sigma_A} < \frac{X-\mu_A}{\sigma_A} < \frac{60.5-\mu_A}{\sigma_A} = \frac{39.5-70}{4.58} < Z < \frac{60.5-70}{4.58}$$

$$= -6.66 < Z < -2.07$$

$$\beta = P(-6.66 < Z < -2.07) = P(Z < +6.66) - P(Z < +2.07) = \boxed{0.0192}$$

KEY TABLE

	H_0 true	H_A true
IDC says not reject H_0 from data		"False Accept H_0" "Type II error" Prob $= \beta$
IDC says reject H_0 from data	"False Reject" "Type I error" Prob $= \alpha$	prob $= 1-\beta$ "POWER of test"

← Fail to reject H_0 against H_A

"If H_A is true, the prob of detecting H_A true with this IDC"

For the problem at hand:

	0.0192
0.0358	0.981

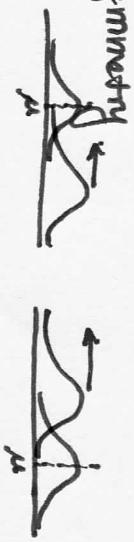
see probs ⑪ and ⑫ for more →

Schuams ch 10

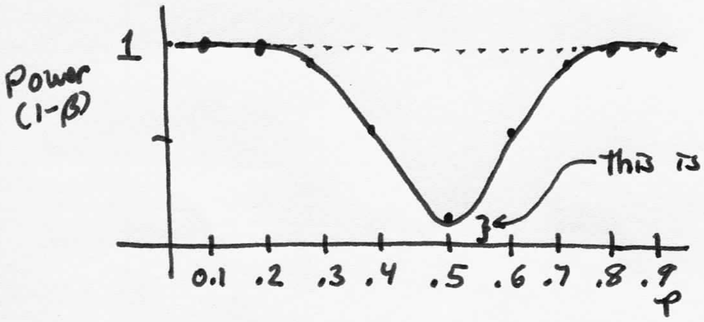
(11) & (12) we can repeat the prev problem for H_A as a fcn of p and see how the power varies. This is an important aspect of the analysis.

p	N parameters	Calculated interval	$\beta = P(a < Z < b)$
0.1			0.0
0.2			0.0
0.3			0.0192
0.4	$\mu_{0.4} = 100(.4) = 40$ $\sigma_{0.4} = \sqrt{100(.6)(.4)} = 4.9$	$\frac{39.5-40}{4.9} < Z < \frac{60.5-40}{4.9}$ $-0.102 < Z < 4.18$	0.504
0.5	This was H_0 $\mu_{0.5} = 50$ $\sigma_{0.5} = 5$	$\frac{39.5-50}{5} < Z < \frac{60.5-50}{5}$ This is exactly the non-tails area of H_0 bump	$\beta = 0.9642$ (Tails area = $1 - 0.9642 = 0.0358$)
0.6	$\mu_{0.6} = 60$ $\sigma_{0.6} = 4.9$	$\frac{39.5-60}{4.9} < Z < \frac{60.5-60}{4.9}$ $-4.18 < Z < 0.102$	0.504
0.7	$\mu_{0.7} = 70$ $\sigma_{0.7} = 4.58$	This is prob 10 $-6.66 < Z \; -2.07$	0.0192
0.8			0.0
0.9			0.0

Note the symmetry
This is because



(12) If we graph Power = $(1-\beta)$ as a fcn of p, we see the regions where there is a high probability of detecting H_A is true, and where there is low probability — where this test (i.e. IDC) is weak.



The test (IDC cutoff value) is powerful when $H_A(p)$ has p far away from $p = 1/2$ (which is H_0)

Power drops to a min at $p = 1/2$. There we are just testing H_0 alone.

(13) we want to detect a significant difference in the mean breaking strength of ropes. Old ropes have mean breaking strength 300 lbs and std dev 24 lbs. Are new ropes better? Design test using $n = 64$ ropes.

(a) Testing using Fisher's methodology

Ingredients: RV $X :=$ weight where rope breaks

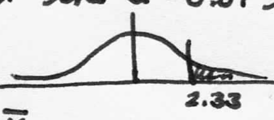
- $\bar{X} :=$ avg over sample of $n = 64$ $\quad \bar{X} \sim f_{\bar{X}} \approx \mathcal{N}(\mu, \frac{\sigma^2}{n})$ — are they give us population params $\mu = 300$, $\sigma = 24$
- H_0: there is no difference i.e. $\bar{X} = 300$
  If we regard the 64 new ropes as a sample from the population of old ropes, would it be an outlier? $H_0 =$ no
  [There is another way to interpret this in prob 17 which I will show also]
- set level of signif $\alpha = 0.01$
  we do a 1 tail test since we are only interested in $\mu < \bar{X}$.

setup for $\mathcal{N}$ approx

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sqrt{\operatorname{var}(\bar{X})}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X} - 300}{24/\sqrt{64}} = \frac{\bar{X} - 300}{3}$$

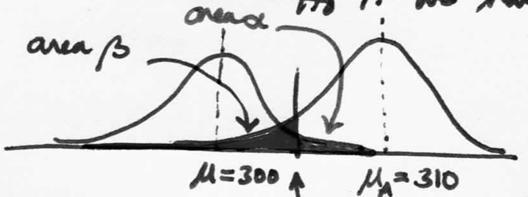
what cond on $\bar{X}$ would cause us to reject H_0 at the $\alpha = 0.01$ level?

$\alpha = 0.01 \Rightarrow Z = 2.33$ for 1-tail test [Problem 5b]



$2.33 \le \frac{\bar{X} - 300}{3} \Rightarrow$ reject H_0 if $307 < \bar{X}$

(b) Now employ Neyman-Pearson — we can calculate the prob of "false acceptance" of H_0 if we have a specific alternative H_A



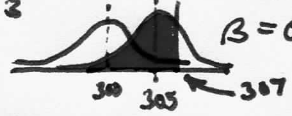
observe that areas α and β can never overlap

Setup $\bar{X}$ is RV again
Distrib family $\mathcal{N}$ same
H_0 same $\mu = 300$
H_A: $\mu_A = 310$ and thus $\bar{X} \sim \mathcal{N}(310, \frac{(24)^2}{64})$
If H_A would be true, $Z_\beta = \frac{\bar{X} - \mu_A}{\sigma_A/\sqrt{n}} \stackrel{!}{\le} \frac{307 - 310}{24/\sqrt{64}} = \frac{-3}{3} = -1$

$\beta = P(Z \le -1) = 0.1587$ from table. This is prob of false accept H_0 or type II error
$= 1 - P(Z < 1)$
$= 1 - \frac{1}{2} - Q(1)$

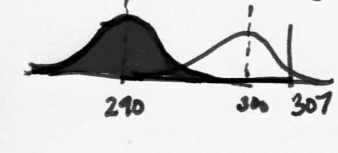
(14) Now Neyman-P extend it to a range of μ_A values [H_A is parameterized by μ_A; we assume $\sigma_A = \sigma$ for all]

$\mu_A = 305$ $\quad Z = \frac{307 - 305}{3} = 0.667 \rightarrow P(Z < 0.67) = \frac{1}{2} + Q(0.67) = 0.7486$



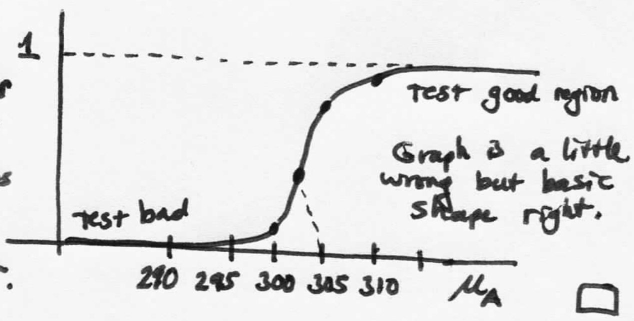
$\beta = 0.7486$

$\mu_A = 290$ $\quad Z = \frac{307 - 290}{3} = 5.67$



$\beta = 1$

This graph characterizes this test and how well it correctly detects when new rope is stronger.



Schuams ch 10 10/31/2019

15

Here we want to show that we can choose arb small values for error probabilities α and β $[0<\alpha,\beta<1]$ and then by making sample size n large enough, we can attain them.

Schuam's problem is a specific example with $\alpha = 0.05 = \beta$. Then I will generalize.

▷ Coin flipping again: $X = \#$ of heads

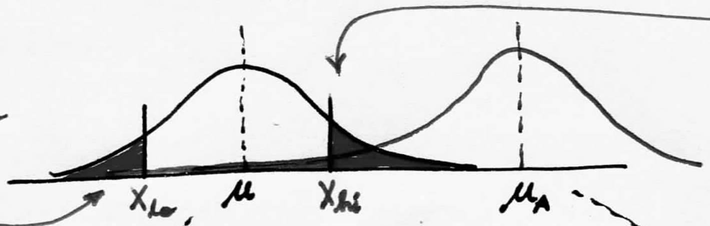
$X \sim Bin(N, p)$ where we will vary N (I used capital N here for number of trials)

Here we choose 2-tail

$H_0: p = \frac{1}{2} = 0.5 \Rightarrow \mu = Np = 0.5N$ $\quad \sigma = \sqrt{Np(1-p)} = 0.5\sqrt{N}$

$H_A: p = 0.6 \Rightarrow \mu_A = 0.6N$ $\quad \sigma_A = 0.49\sqrt{N}$

We are assuming any green area here B so small as to be negligible



we want to show such a value exists that satisfies the requirements (does not have to be an integer).

Normalize

$$Z = \frac{X - E(X)}{\sqrt{var(X)}} = \frac{X-\mu}{\sigma} = \frac{X - \frac{1}{2}N}{\frac{1}{2}\sqrt{N}}$$

From the picture, we see we want $Z_\alpha < Z$ [really $Z_{\alpha/2}$ since 2 tail but I omit that in notation]

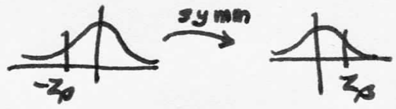
Also for 2-tail and $\alpha = 0.05$, we know $Z_\alpha = 1.96$ but plug that in later.

$$Z_\alpha < \frac{X - \frac{1}{2}N}{\frac{1}{2}\sqrt{N}}$$

$$\Rightarrow \frac{1}{2}N + \frac{Z_\alpha}{2}\sqrt{N} < X$$

Here $Z = \dfrac{X - \mu_A}{\sigma_A}$

we have 1 sided test



For $\beta = 0.05$, $Z_\beta = 1.645$

$$\frac{X - 0.6N}{.49\sqrt{N}} < -Z_\beta$$

$$X < 0.6N - Z_\beta(0.49)\sqrt{N}$$

combine these to show cutoff X exists

$$\frac{1}{2}N + \frac{Z_\alpha}{2}\sqrt{N} < X < 0.6N - Z_\beta(0.49)\sqrt{N}$$

$$\Rightarrow \frac{1}{2}N + \frac{Z_\alpha}{2}\sqrt{N} < 0.6N - Z_\beta(0.49)\sqrt{N}$$

$$\left(\frac{Z_\alpha}{2} + 0.49Z_\beta\right)\sqrt{N} < (0.6 - 0.5)N$$

$$\Rightarrow \frac{\left(\frac{Z_\alpha}{2} + 0.49Z_\beta\right)}{(0.1)} < \sqrt{N} \Rightarrow 100\left(0.5Z_\alpha + 0.49Z_\beta\right)^2 < N$$

plug in values

$$100\left(\frac{1.96}{2} + 0.49(1.645)\right)^2 < N$$

$$318.997 < N$$

Take $N = 319$ or larger.

Then $0.5(319) + \frac{1.96}{2}\sqrt{319} < X < 0.6(319) - 1.645(0.49)\sqrt{319}$

177.0033597 $\qquad$ 177.0034867

so such a cutoff X can exist between them.

cont'd →

10/31/2019

15 cont'd

Now extend it: More generally we would have

really $\alpha/2$ again

$$Z_\alpha < \frac{X-\mu(N)}{\sigma(N)} \quad \text{and} \quad \frac{X-\mu_A(N)}{\sigma_A(N)} < -Z_\beta$$ Z_β itself always pos

$$\mu(N) + Z_\alpha\sigma(N) < X < \mu_A(N) - Z_\beta\sigma_A(N)$$

To go further, we need an explicit form for $\mu(N)$ and $\sigma(N)$ where we see N dependence. Here we get it from assuming $\mathcal{N}$ approx to <u>Binomial</u> distrib.

we can be more general about p ~~value~~ too: p_0 vs. p_A

$$Np_0 + Z_\alpha\sqrt{Np_0(1-p_0)} \overset{?}{<} Np_A - Z_\beta\sqrt{Np_A(1-p_A)}$$

$$Z_\alpha\sqrt{p_0(1-p_0)}\sqrt{N} + Z_\beta\sqrt{p_A(1-p_A)}\sqrt{N} < N(p_0-p_A)$$

$$\boxed{\left[\frac{Z_\alpha\sqrt{p(1-p)} + Z_\beta\sqrt{p_A(1-p_A)}}{(p_A-p_0)}\right]^2 < N}$$ This can always be satisfied, unless $p_A = p_0$

Now REA Statistics p. 642 considers a similar prob, but with the <u>approx by $\mathcal{N}$</u> sampling distrib for $\bar{X}$ ($f_{\bar{X}}$) rather than binomial. But again since $f_{\bar{X}} \sim \mathcal{N}(\mu, \sigma/\sqrt{n})$

<u>Thm</u> Given preset α, β values

$H_0: \mu = \mu_0$ and σ } same

$H_A: \mu = \mu_A$ and σ

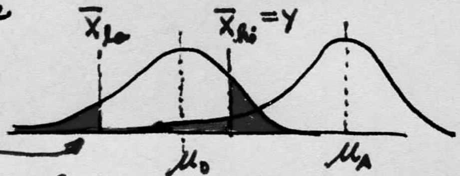
$\mu_0 \neq \mu_A$

$\Longrightarrow$ we can always meet these requirements with sample size

$$n > \frac{(Z_\alpha + Z_\beta)^2\sigma^2}{(\mu_A-\mu_0)^2}$$

pf. very similar to before

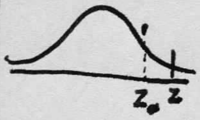
Green spill over area assumed negligible



We normalize the H_0 pdf

$Z := \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$

Since here we happen to be considering a 2-tail test, there are cutoff values $\bar{X}_{lo}$ and $\bar{X}_{hi}$ that demarcate the α tails. Here $\mu_0 < \mu_A$ so we are interested in $\bar{X}_{hi}$ (= Y for brevity of notation)



Thus $Z_\alpha \le \frac{Y-\mu_0}{\sigma/\sqrt{n}} \Rightarrow Z_\alpha\frac{\sigma}{\sqrt{n}} + \mu_0 \le Y$ [where as usual Z_α means $Z_{\alpha/2}$]

Likewise for normalizing the H_A pdf $Z := \frac{\bar{X}-\mu_A}{\sigma/\sqrt{n}}$

$\Rightarrow \frac{Y-\mu_A}{\sigma/\sqrt{n}} \le -Z_\beta$

$\beta := P(Z < -Z_\beta)$ but this would be more complex to compute if we had to subtract green area to the left of $\bar{X}_{lo}$ so we assume it is negligible.

$\triangleright$ Then same as before: $Z_\alpha\frac{\sigma}{\sqrt{n}} + \mu_0 \le Y \le -Z_\beta\frac{\sigma}{\sqrt{n}} + \mu_A$

$$\Rightarrow Z_\alpha\frac{\sigma}{\sqrt{n}} + Z_\beta\frac{\sigma}{\sqrt{n}} \le \mu_A - \mu_0 \Rightarrow \frac{(Z_\alpha+Z_\beta)^2\sigma^2}{(\mu_A-\mu_0)^2} \le n$$

There is always an n satisfying this

QED

Schuams ch 10 11/1/2019

(16)

Here we show the close relationship between Confidence Intervals and Hypothesis Tests

A machine makes BBs. $\mu = 0.574$ inch diam, $\sigma = 0.008$. Take a sample of $n=6$ every hour and measure them for quality control.

How can we design a decision rule to sound the alarm if the machine has gone out of alignment?

- RV X = diam of BB; we want $\bar{X}$ = avg diam of sample
- Distrib $f_{\bar{X}} \approx \mathcal{N}(\mu_{\bar{X}}, \sigma_{\bar{X}}^2)$
- H_0 : machine is working fine.
- Take $\alpha = 0.01$

Following my sheet for prob #13(a), we assume the $\bar{x}$ for the sample came from H_0 and we want to know if it is an outlier (so improbable that we reject H_0)

For $\alpha = 0.01$, we get $z_c = 2.58$

We form

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{var(\bar{X})}} = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

and must have $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < -2.58$ or $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} > 2.58$ TO REJECT H_0.

Convert this to a CI:

H_0 is <u>not</u> rejected if $\underset{-z_c}{-2.58} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \underset{z_c}{2.58}$

$$\Rightarrow -z_c\frac{\sigma}{\sqrt{n}} + \mu < \bar{X} < z_c\frac{\sigma}{\sqrt{n}} + \mu$$

99% CI

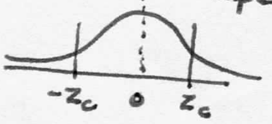
$$\Rightarrow \bar{X} \in \left(\mu - z_c\frac{\sigma}{\sqrt{n}}, \ \mu + z_c\frac{\sigma}{\sqrt{n}}\right)$$

So if $\bar{X}$ lies outside this range, sound the alarm.

$$= \left(0.574 - \frac{2.58(0.008)}{\sqrt{6}}, \ 0.574 + \frac{2.58(0.008)}{\sqrt{6}}\right)$$

$\square$

<u>More generally</u>, for CI we have an unknown pop μ and we compute $\bar{X}$ from sample size n. We assume $f_{\bar{X}} \approx \mathcal{N}(\mu, \frac{\sigma^2}{n})$. Lets say we want 95% for μ. so we form $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ and must have $-z_c < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_c$



un normalize

$$\Rightarrow \bar{X} - z_c\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_c\frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow \mu \in \left(\bar{X} - z_c\frac{\sigma}{\sqrt{n}}, \ \bar{X} + z_c\frac{\sigma}{\sqrt{n}}\right)$$

For Hypoth Testing (for the mean), again we form $Z = \frac{\bar{X}-\mu}{\sigma_{\bar{X}}}$

we "accept" H_0 if $Z \in (-z_c, z_c)$ for 2 tail test

unnormalize $-z_c < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_c$ <u>but</u> here we solve for $\bar{X}$ rather than μ

$$\bar{X} \in \left(\mu - z_c\frac{\sigma}{\sqrt{n}}, \ \mu + z_c\frac{\sigma}{\sqrt{n}}\right)$$

observe these are not the same interval, unless we are lucky enough that $\bar{X} = \mu$ $\square$

11/1/2019

(17) (Reformulated) $\boxed{\text{We want to detect the difference between 2 populations based on samples}}$

Is there a difference in intelligence between boys and girls?
Choose a sample of $n=40$ girls and give test [X_i = score of i^{th} girl]
Give same test to $m=50$ boys [Y_i = score of boy i]
From the samples $\bar{x}=74$ $\bar{y}=78$
$s_{\bar{x}}=8$ $s_{\bar{y}}=7$
Does this indicate a significant difference between the populations? i.e $\mu_x \neq \mu_y$

Fisher setup RV: $(\bar{X}-\bar{Y})$
Distrib: $(\bar{X}-\bar{Y}) \sim f_{\bar{X}-\bar{Y}} \approx N(\mu_{\bar{X}-\bar{Y}}, \sigma^2_{\bar{X}-\bar{Y}})$
H_0: $\mu_x = \mu_y$ (against $\mu_x \neq \mu_y$, so 2-tail test)

Setup for N approx
$$Z_0 = \frac{(\bar{X}-\bar{Y}) - E(\bar{X}-\bar{Y})}{\sqrt{Var(\bar{X}-\bar{Y})}}$$

Now $E(\bar{X}-\bar{Y}) = E(\bar{X}) - E(\bar{Y}) = E(X) - E(Y)$
$= \mu_x - \mu_y = 0$ by H_0

$$\begin{aligned} Var(\bar{X}-\bar{Y}) &= Var(\bar{X}) + Var(\bar{Y}) \\ &= Var\left(\tfrac{1}{n}\textstyle\sum X_i\right) + Var\left(\tfrac{1}{m}\textstyle\sum Y_i\right) \\ &= \tfrac{1}{n^2} Var\left(\textstyle\sum X_i\right) + \tfrac{1}{m^2} Var\left(\textstyle\sum Y_i\right) \\ \text{X,Y indep} \quad &= \tfrac{1}{n^2} \textstyle\sum Var(X_i) + \tfrac{1}{m^2} \textstyle\sum Var(Y_i) \\ &= \tfrac{1}{n} s_x^2 + \tfrac{1}{m} s_y^2 \end{aligned}$$

$$\Rightarrow \sigma_{\bar{X}-\bar{Y}} = \sqrt{\tfrac{1}{40} 8^2 + \tfrac{1}{50} 7^2} = 1.606$$

$$\Rightarrow Z_0 = \frac{(\bar{X}-\bar{Y}) - 0}{\sigma_{\bar{X}-\bar{Y}}} = \frac{(74-78)-0}{1.606} = -2.49$$

Compute p-Value

$\boxed{\begin{array}{c} Z=-2.49 \\ \text{by symm, we can} \\ \text{work with } Z=2.49 \end{array}} \xrightarrow{\text{reverse table lookup}} \boxed{\begin{array}{c} \frac{1}{2} - \hat{\alpha} \overset{!}{=} 0.4936 \\ \hat{\alpha} = 0.0064 \end{array}} \xrightarrow{\text{This is 2-tail so } \alpha = 2\hat{\alpha}} \boxed{\alpha = 0.0128}$

This is the p-value

Thus we would have $0.01 < \text{p-Value} < 0.05$
we would reject H_0 at 0.05 level, but not at 0.01

$\square$

8/25/2019

(18) again consider 2 populations. Is there a significant difference in the mean value of a certain characteristic between them?
[The difference between this prob and (17) is now we also consider 1-tailed test]

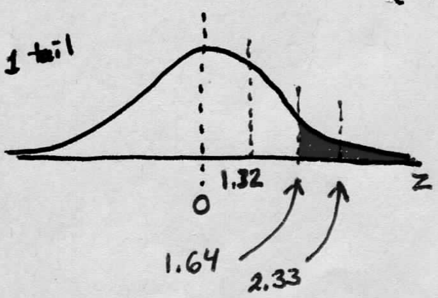
		sample		
Studs who like sports	$\sim X :=$ height of stud	$n = 50$	$\bar{x} = 68.2$	$S_{\bar{x}} = 2.5$
Studs who like books	$\sim Y :=$ height of stud	$m = 50$	$\bar{y} = 67.5$	$S_{\bar{y}} = 2.8$

$H_0 : \mu_x = \mu_y$ compared against $H_A : \mu_x > \mu_y$ i.e. $\mu_x - \mu_y > 0$

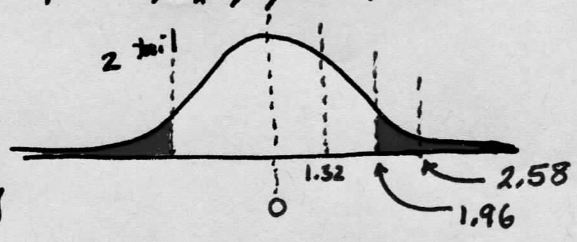
As usual, first we normalize:

$$Z_0 = \frac{(\bar{x}-\bar{y}) - E(\bar{x}-\bar{y})}{\sqrt{Var(\bar{x}-\bar{y})}} = \frac{(\bar{x}-\bar{y}) - 0}{\sqrt{\frac{1}{50}S_x^2 + \frac{1}{50}S_y^2}} = \frac{(\bar{x}-\bar{y})-0}{0.53} = \frac{(68.2-67.5)-0}{0.53} = 1.32$$

Can we reject H_0 at significance level α? Lets compare different α values as well as 1 and 2 tail tests (book considers only 1 tail, since $\mu_x > \mu_y$ is H_A). Also compute p-Value.



area = probability



$\alpha = 0.05 \xrightarrow{\text{1 tail}} \left[\frac{1}{2} - 0.05 = 0.45\right] \longrightarrow [0.4500] \xrightarrow{\text{table}} 1.64$ $1.64 > 1.32$ so can't reject H_0

$\xrightarrow{\text{2 tail}}$ $\tilde{\alpha} = \frac{1}{2}\alpha \rightarrow \left[\frac{1}{2} - 0.025 = 0.4750\right] \rightarrow [0.4750] \rightarrow 1.96$ can't reject

$\alpha = 0.01 \xrightarrow{\text{1 tail}} \left[\frac{1}{2} - 0.01 = 0.49\right] \rightarrow [0.4900] \xrightarrow{\text{table}} 2.33$ can't reject

2 tail $\left\{ \tilde{\alpha} = \frac{1}{2}\alpha \rightarrow \left[\frac{1}{2} - 0.005 = 0.495\right] \rightarrow [0.4950] \rightarrow 2.58 \right.$ can't reject

<u>p-Value</u>

$\alpha = \tilde{\alpha}$

$\alpha = 0.0934$ this is p-Value $\xleftarrow{\text{1 tail}}$ $\left[\frac{1}{2} - \tilde{\alpha} \overset{!}{=} 0.4066 \Rightarrow \tilde{\alpha} = 0.0934\right] \longleftarrow [0.4066] \xleftarrow{\text{reverse table}} Z_0 = 1.32$

$\alpha = 0.1868$ p-Value $\xleftarrow{\text{2 tail}}$ $\alpha = 2\tilde{\alpha}$

We could reject H_0 if we accept prob of false reject 0.1868 or higher $\Rightarrow$ approx 19%

(19) How much should we increase the sample size n to be able to reject H_0?

Here we set $\alpha = 0.05$

So we want
$$Z_0 = \frac{(\bar{x}-\bar{y}) - E(\bar{x}-\bar{y})}{\sqrt{Var(\bar{x}-\bar{y})}} = \frac{(\bar{x}-\bar{y})-0}{\frac{1}{\sqrt{n}}\sqrt{S_{\bar{x}}^2 + S_{\bar{y}}^2}} \overset{!}{>} 1.64$$

$$= \frac{0.7}{\frac{1}{\sqrt{n}}(3.754)} > 1.64$$

$$\Rightarrow \sqrt{n} > 8.795 \Rightarrow n > 77.35 \text{ or } n > 78$$

This is a bit fake because it is saying $\bar{x}(n) = 68.2$ <u>for all values of n</u>. And likewise for $\bar{y}$, $S_{\bar{x}}$, $S_{\bar{y}}$

□