Chapter 9 | Hypothesis Tests

Intro

9.1

to follow on next page

In this chapter we utilize the Central Limit Theorem in a different approach than we did in Confidence Intervals (chpt 8).

- We test people's claims in a hypothesis Test:

ex The Signal, Santa Clarita's Newspaper, claims that 64% of Santa Claritans favor the Death Penalty.

- Perhaps you doubt that claim. So we will test it. This is called Hypothesis Testing.

- In logic the conditional statement has the form: "If A then B"

ex If "you water the grass" then "it will be green"

The 1st statement in this conditional is " you water ~~the~~ grass" It is called the hypothesis of the test.

A = hypothesis : water the grass

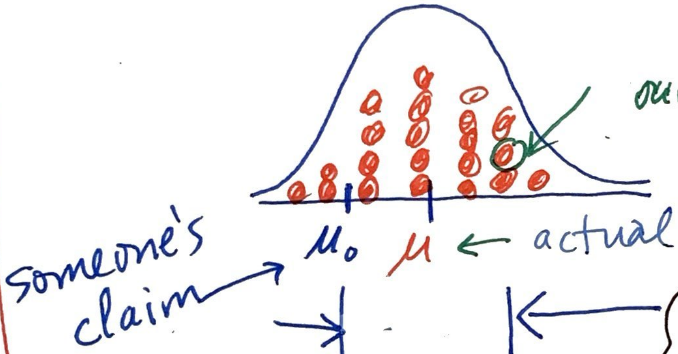
B = conclusion : will be green

②

- We will collect our own sample... but we will expect it to vary from "the claim"s

We ask the question: Is this descrepency we see due to sample variablity OR is the descrepany due to statistical significance?

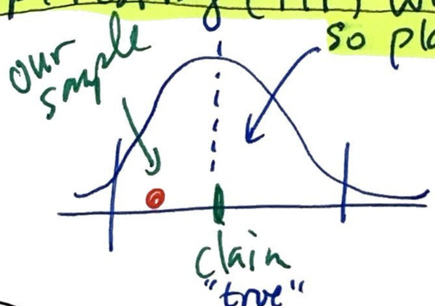
- Consider the "Gods eye view" of a population parameter



Is this distance "sample variability" OR "statistical significance"?

- In Hyp. Testing (HT) we assume the claim to be true so place it under the center of a unimodal dist

Findings

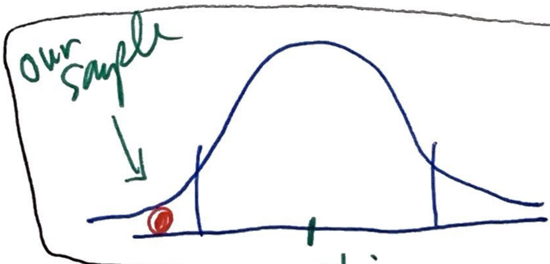


= statistical variability

~~"we agree"~~ don't say

instead say "we can't disagree"

or better... "we fail to reject your claim"



= statistical significance

"we reject your claim"

or say "very unlikely" we now doubt your claim

Big Picture for Hypothesis Testing

In this chapter we test a claim that someone makes regarding a population parameter

we question their results and gather our own data to see if out data supports, or, not supports, their claim

We treat our analysis like a trial by jury:
"Innocent until proven guilty"
- So that means we believe their claim until we are shown otherwise.
- IF we do have significant deviation from their claim - we can reject their claim {guilty beyond a reasonable doubt}
- If our data is close enough to their claim we can support the claim with some conditions:
  - we have normal statistical variation

we say "we fail to reject" your claim.

'lt

- Unlike a trial we can pin a statistical value to the "deviation" we see between our data and the claim.

- This statistical value is called the p-value.

"The probability our data could occur given that their claim is correct"

⊛ The 4 major steps followed in a Hyp. Test:

After we Gather our own data (poll or research)

1. State the hypotheses
2. Establish which model to use
3. Do the mechanics (do the math)
4. State the conclusion

2

Conditions to be met…

STEP Ø

Proportions (z)

One sample

Conditions	Justification
1. Individuals are independent.	1. SRS and $n < 10\%$ of the population.
2. Sample is sufficiently large.	2. Successes and failures each ≥ 10.

Two Groups

1. Groups are independent.	1. (Think about how the data were collected.)
2. Data in each group are independent.	2. Both are SRSs and $n < 10\%$ of populations OR random allocation.
3. Both samples are sufficiently large.	3. Successes and failures each ≥ 10 for both groups.

Means (t)

One Sample (df = $n - 1$)

1. Individuals are independent.	1. SRS and $n < 10\%$ of the population.
2. Population has a Normal model.	2. Histogram is unimodal and symmetric.*

Matched pairs (df = $n - 1$)

1. Data are matched.	1. (Think about the design.)
2. Individuals are independent.	2. SRS and $n < 10\%$ OR random allocation.
3. Population of differences is Normal.	3. Histogram of differences is unimodal and symmetric.* or n>30

Two independent samples (df from technology)

1. Groups are independent.	1. (Think about the design.)
2. Data in each group are independent.	2. SRSs and $n < 10\%$ OR random allocation.
3. Both populations are Normal.	3. Both histograms are unimodal and symmetric.* or both n>30

Distributions/Association (χ^2)

Goodness of fit (df = # of cells − 1; one variable, one sample compared with population model)

1. Data are counts.	1. (Are they?)
2. Data in sample are independent.	2. SRS and $n < 10\%$ of the population.
3. Sample is sufficiently large.	3. All expected counts ≥ 5.

Homogeneity [df = $(r-1)(c-1)$; many groups compared on one variable]

1. Data are counts.	1. (Are they?)
2. Data in groups are independent.	2. SRSs and $n < 10\%$ OR random allocation.
3. Groups are sufficiently large.	3. All expected counts ≥ 5.

Independence [df = $(r-1)(c-1)$; sample from one population classified on two variables]

1. Data are counts.	1. (Are they?)
2. Data are independent.	2. SRSs and $n < 10\%$ of the population.
3. Sample is sufficiently large.	3. All expected counts ≥ 5.

Regression (t, df = $n - 2$)

Association of each quantitative variable ($\beta = 0$?)

1. Form of relationship is linear.	1. Scatterplot looks approximately linear.
2. Errors are independent.	2. No apparent pattern in residuals plot.
3. Variability of errors is constant.	3. Residuals plot has consistent spread.
4. Errors have a Normal model.	4. Histogram of residuals is approximately unimodal and symmetric, or Normal probability plot reasonably straight.*

(*less critical as n increases)

We now march through the steps

1. Stating Hypotheses

1st: our starting hypothesis is the null hypoth.

- denoted by $\boxed{H_0}$
- assumed to be a pop. parameter. So we use pop. symbology like μ, σ {vs $\bar{x}, s$}
  - pop. parameters: p, μ, σ

the null hyp. is assumed to be the "status quo." "presently accepted value"

- we write down the null hyp. in the form $\boxed{H_0 : \text{param} = \text{hypothesized valu.}}$

ex $\boxed{H_0 : p = 0.64}$

the claim that Santa claritan's have a 64% favor of the death penalty.

2^{nd}: we also form the alternative hypothesis. denoted as H_A { or H_1 }.

- H_A contains the values of the pop. param that we consider plausible should we reject the null hypothesis.

{ what is our criteria for rejection? }

- There are 3 possible alt. hypotheses.

	we think	vs. the claim
1.	H_A: pop. param <	the hypothesized claim.
2.	H_A: pop. param $\neq$	hyp. claim.
3.	H_A: pop. param >	hyp. claim

ex) $\boxed{H_A : p \neq 0.64}$ we think the claim of 0.64 is too low or too high

Conclude:

- We want to compare our data to what we would expect given that H_0 is true.
- we can do this by finding out how many std. dev. our data is from your claim (a.k.a. the test statistic).

" Q: How likely is it to get our result if their claim, H_0, is true?

2 Establishing the Model

- We need to specify the CLThm model to use to test the claim, Ho.

  N (μ, SD) or Z-table
  t-table
  χ^2-table

- All models require a set of assumptions, or conditions, to be valid. (step 0)

- we state the condition and what we think justifies its satisfaction.

> * if we cannot meet the conditions then the testing stops.

- Each test has a name that should be included in our repsonse (report)

Ex The test for proportion of Santa Claritians who support the death penalty require a "one-proportion Z-test"

(step 0) conditions to be met for this test are:

- randomization
- 10% condition (Not too many to render our sample dependent)
- Success & Failures both larger that 10. to use the Central Limit Thm.

3 the mechanics

- we perform the actual calculations to obtain the "test statistic" for our data.

mean, σ known — z-score
mean, σ unknown — t-score
proportions — z-score

- Different models (tests) have different formulas and different test statistics.
- Use the worksheet and/or software

⊗ For the "p-value method" the ultimate goal of the calculation is to obtain a "P-value"

- the P-value is the probability that the observed statistic from our data could occur assuming the null model, the claim, is indeed the correct value.

- If the p-value is small enough then we reject the null hyp. (the claim).

The p-value is a conditional probability

P-value = P(our test data could occur | their claim is true)

"given that"

3 The mechanics (continued)

EX For one-proportion z-test regarding the percentage of Santa Claritans prefering the Death Penalty

- conditions (were discussed previously)

- H_0 : $p = 0.64$
- test statistic : $z = \dfrac{(\hat{p} - p_0)}{SD}$ our sample; claim of the pop.

  - $SD = \sqrt{\dfrac{p_0 q_0}{n}}$ pop. prop. of successes; pop. prop. of failures; sample size

{ we use pop. SD vs. SE since we assume the claim is refering to the population }

{ Once these values are calculated we state that we will use the

$$N(Z, SD)$$
}

- calculate the z-score for your

$$\boxed{Z = \frac{\hat{P} - P_0}{SD}}$$

assume this is known.

we go off to the z-table to get the p-value

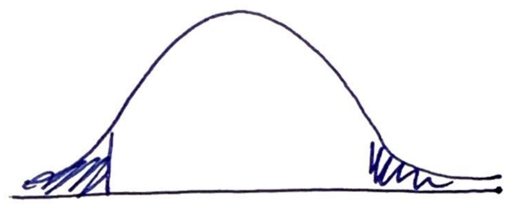
So we need to areas uner choosen distribution in the left, the right tails, or both tails.

$\boxed{\text{Which of these configurations we obtain depends on } H_A \text{ \{the hypothesis we are using to test the claim\}}}$

Some Details

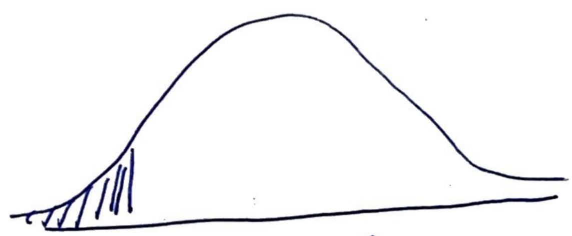
choices

$H_A : p \neq p_0$



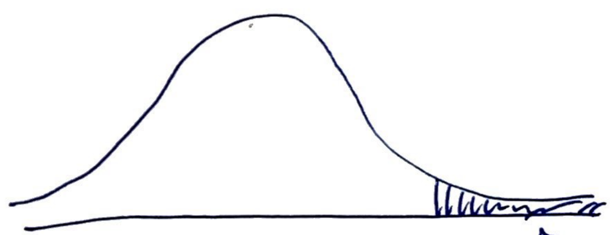
p-value = area comb. in both tails

$H_A : p < p_0$



p-value = area

$H_A : p > p_0$



p-value = area

Once we have the area(s), the p-value, we move into the next step...

4 State the Conclusion

the conclusion in a hypothesis test is always a statement about the null hyp.'s {i.e. the claim} validity.

- we state either that "we reject the null"

  -OR-

  "we fail to reject the null hypothesis"

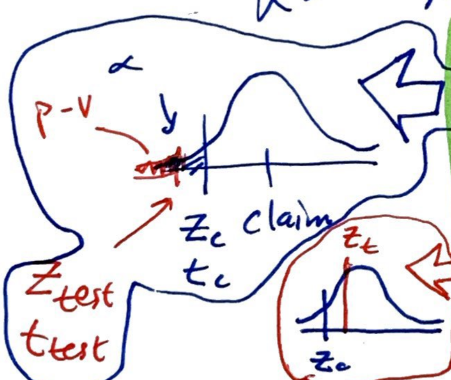
  "do not accept"

- state the conclusion in context.

So if p-value $\lesseqgtr$ α value

α-value is the demarcation of when you decide that a sample is statistically significant or statistical variation

$\alpha \equiv$ the provided significance level



$p < \alpha$	reject the claim	statistically significant
$p > \alpha$	fail to reject the claim	statistical variation

- state that "given the claim is true then our data is fails to reject the claim"

  or "we reject the claim ..."

The Flagship Example

EX

Historically (pre 2000) the percentage of U.S. residents who support stricter gun control laws has been 52%. A recent Gallup Poll of 1011 people showed 495 in favor of stricter gun control laws. Assume that the poll was a random sample.

Q: Test the claim that the proportion of those favoring stricter laws has changed from 0.52 by performing a hyp. test.

(Assume a significance level of $\alpha = 0.05$)

- choose a one proportion z-test

Conditions

- Random sample: stated
- is the sample > 10% of the pop.?
  i.e. is the pop. more than 10 * 1011? 10,110 yes!
- 10 successes / 10 failures

$$np_0 = (1011)(0.52) = 526 \text{ in favor} > 10$$

$$n(1-p_0) = (1011)(0.48) = 485 \text{ against} > 10$$

[EX] (cont.)

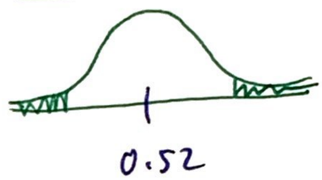
- Hypotheses

H_0 : the pop. proportion remains @ 0.52 who support stricter gun laws.

H_A : the propoportion is no longer at 0.52 {it is either more or less}

i.e.

$$\boxed{\begin{aligned} H_0 &: P_0 = 0.52 \\ H_A &: P_0 \neq 0.52 \end{aligned}}$$ → So use a two-tails test

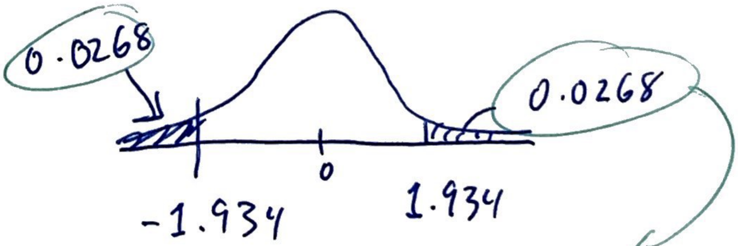


- use $\alpha = 0.05$ (requested)
- formulas $SD = \sqrt{\frac{P_0 q_0}{n}} = \sqrt{\frac{(0.52)(0.48)}{1011}} = \underline{\underline{0.0157}}$

∴ $\hat{p} = 495/1011 = 0.4896 = \underline{\underline{0.490}}$

∴ $z = \frac{0.490 - 0.52}{0.0157} = \underline{\underline{-1.934}}$

$P(z < -1.934) = 0.0268$



add to gether

∴ $\boxed{\text{P-Value} = 0.0536}$

Ex (Cont.)

- our $\boxed{\text{p-value of 0.0536 is above}}$ the decided significance level of $\alpha = 0.05$

- we "fail to reject" the claim that U.S. residence proportion of 0.52 wanting stricter gun laws has changed.

Conclude:
"There is insufficient evidence to conclude that the proportion is different than the claim of 52% who favor stricter gun laws"

statdisk.com

Analysis → Hyp-Testing → Prop'n One Sample

1) Pop. proportion ≠ Claimed ⌄ ← keep.

- Sig. $\alpha = \boxed{0.05}$
- Claimed proportion : $p = \boxed{0.52}$
- Sample size : $n = \boxed{1011}$
- Successes : $x = \boxed{495}$

Evaluate

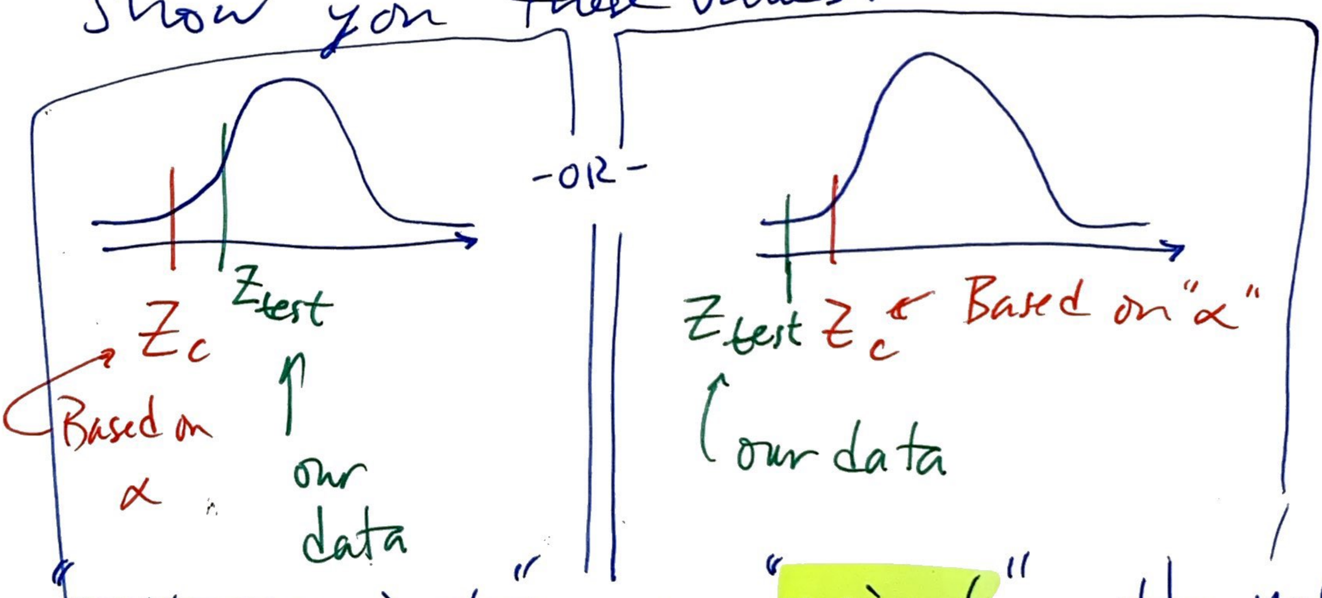
output $\boxed{\text{pvalue: } 0.05313}$ greater than $\alpha = 0.05$ ⇒ "fail to reject"

⊛ Alternative Method: "critical value" method.

- Instead of finding the p-value from the test statistic { z_{test}, t_{test} }, compare the z_{test} (t_{test}) to the critical values z_c (t_c) that are obtained from the tables.

See the last page: look up, under the two-tail header, $\alpha = 0.05$ and zoom to the bottom to get z_c

- If $|z_{test}| > |z_c|$ we reject
  If $|z_{test}| < |z_c|$ we fail to reject
- statdisk will, when you click on "plot" show you these values:



"Fail to reject" -OR- "reject" the null.

Ex In the previous example the test statistic was

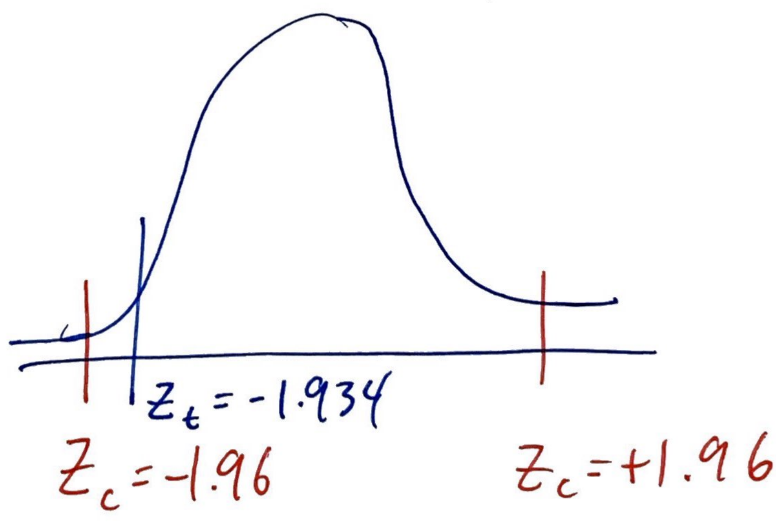
$$Z_t = -1.934$$

From the t-table's bottom line, under the two-tail header (used for $\neq H_A$) in the last line we read

$$Z_{crit} = 1.960$$

Since $|Z_t| < |Z_c|$ we "fail to reject"

{ but it is very close ... go get a few more data to reassess }



↳ see last page for these values

$\boxed{EX}$ (cont.)

statdisk.com :

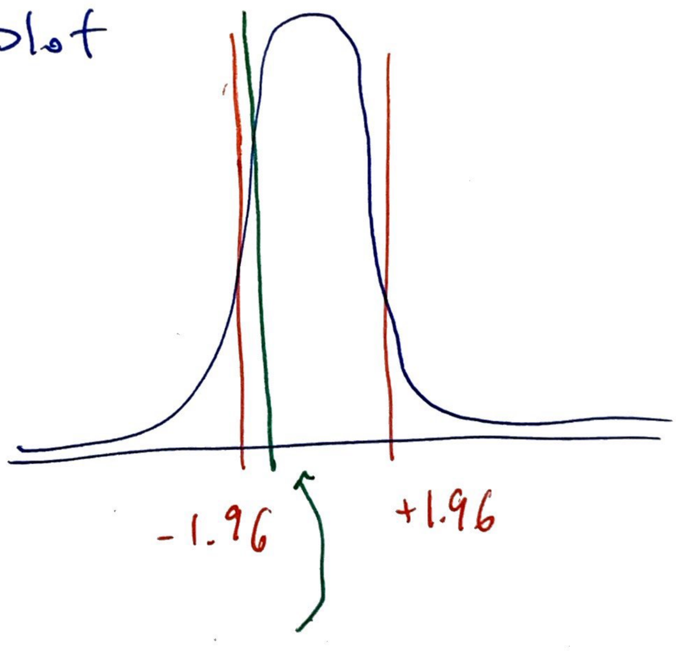
Analysis → H. Test → Prop'n one-sample

$\boxed{1)\ pop \neq claimed}$

- sig $\alpha = \boxed{0.05}$
- claimed : $\boxed{0.52}$
- $n = \boxed{1011}$
- $x = \boxed{495}$

$\boxed{\text{Evaluate}}$

plot



Analysis:

$$Z_{test} = -1.934$$

Z_t lies within $-Z_c < Z < +Z_c$ so "fail to reject"

16

use if $>$ or $<$ one tail test

TABLE A-3 t Distribution: Critical t Values

			Area in One Tail		
	0.005	0.01	0.025	0.05	0.10
			Area in Two Tails		
Degrees of Freedom	0.01	0.02	0.05	0.10	0.20
1	63.657	31.821	12.706	6.314	3.078
2	9.925	6.965	4.303	2.920	1.886
3	5.841	4.541	3.182	2.353	1.638
4	4.604	3.747	2.776	2.132	1.533
5	4.032	3.365	2.571	2.015	1.476
6	3.707	3.143	2.447	1.943	1.440
7	3.499	2.998	2.365	1.895	1.415
8	3.355	2.896	2.306	1.860	1.397
9	3.250	2.821	2.262	1.833	1.383
10	3.169	2.764	2.228	1.812	1.372
11	3.106	2.718	2.201	1.796	1.363
12	3.055	2.681	2.179	1.782	1.356
13	3.012	2.650	2.160	1.771	1.350
14	2.977	2.624	2.145	1.761	1.345
15	2.947	2.602	2.131	1.753	1.341
16	2.921	2.583	2.120	1.746	1.337
17	2.898	2.567	2.110	1.740	1.333
18	2.878	2.552	2.101	1.734	1.330
19	2.861	2.539	2.093	1.729	1.328
20	2.845	2.528	2.086	1.725	1.325
21	2.831	2.518	2.080	1.721	1.323
22	2.819	2.508	2.074	1.717	1.321
23	2.807	2.500	2.069	1.714	1.319
24	2.797	2.492	2.064	1.711	1.318
25	2.787	2.485	2.060	1.708	1.316
26	2.779	2.479	2.056	1.706	1.315
27	2.771	2.473	2.052	1.703	1.314
28	2.763	2.467	2.048	1.701	1.313
29	2.756	2.462	2.045	1.699	1.311
30	2.750	2.457	2.042	1.697	1.310
31	2.744	2.453	2.040	1.696	1.309
32	2.738	2.449	2.037	1.694	1.309
34	2.728	2.441	2.032	1.691	1.307
36	2.719	2.434	2.028	1.688	1.306
38	2.712	2.429	2.024	1.686	1.304
40	2.704	2.423	2.021	1.684	1.303
45	2.690	2.412	2.014	1.679	1.301
50	2.678	2.403	2.009	1.676	1.299
55	2.668	2.396	2.004	1.673	1.297
60	2.660	2.390	2.000	1.671	1.296
65	2.654	2.385	1.997	1.669	1.295
70	2.648	2.381	1.994	1.667	1.294
75	2.643	2.377	1.992	1.665	1.293
80	2.639	2.374	1.990	1.664	1.292
90	2.632	2.368	1.987	1.662	1.291
100	2.626	2.364	1.984	1.660	1.290
200	2.601	2.345	1.972	1.653	1.286
300	2.592	2.339	1.968	1.650	1.284
400	2.588	2.336	1.966	1.649	1.284
500	2.586	2.334	1.965	1.648	1.283
750	2.582	2.331	1.963	1.647	1.283
1000	2.581	2.330	1.962	1.646	1.282
2000	2.578	2.328	1.961	1.646	1.282
Large	2.576	2.326	1.960	1.645	1.282

$\neq$ H. Test

Z_c Z_c