Chat 5
Notes 4 Hypothesis Testing

1. Hypothesis Testing Logic

Summarized:

a. **Form null hypothesis**, e.g., there will be no difference in math scores between males and females;

Ho: $\mu_{males} = \mu_{females}$

b. Collect relevant data from random sample taken defined population
c. Test data relative to null hypothesis – **calculate probability of randomly selecting data like those obtained if the null hypothesis is true**, i.e., check the probability of selected data deviating from what was specified by the null hypothesis assuming the null is true
d. If probability is small, reject Ho; if probability is large fail to reject Ho
e. Provide inference (was Ho rejected) and interpretation (what findings tell us) of results.

2. One Sample Z-test with p-values

One sample Z-test is designed to test whether a sample mean deviates from some hypothesized value.

Note – one sample Z-test is different from Z scores.

- Z scores show have far a raw score deviates from a sample mean in SD units;
- Z-test is used to determine whether a sample mean appears to be different from some set value, and shows how far the sample mean deviates from this value in Standard Error units.

a. Hypotheses

Null

Symbolic → Ho: μ = "some value" (e.g., Ho: μ = 16 oz)
Written → There will be no difference between the sample mean and "some value"

Example Null
Mean number of auto accidents during past 10 years is the same as the mean number of auto accident during the previous 20 years (μ = 24.5) in Statesboro.

Symbolic null → Ho: μ = 24.5
Written null → There will be no difference in mean number of auto accidents between past 10 years and previous 20 years in Statesboro (μ = 24.5, σ = 7.6).

Alternative, Non-directional

Symbolic → Ha: μ ≠ "some value" (e.g., Ha: μ ≠ 16 oz)
Written → There will be a difference between the sample mean and "some value"

Example Non-directional

Symbolic null → Ha: μ ≠ 24.5
Written null → There will be a difference in mean number of auto accidents between past 10 years and previous 20 years in Statesboro (μ = 24.5, σ = 7.6).

Ho: μ = 24.5
H1: μ ≠ 24.5

b. Formula (one sample z-test)

$$\boldsymbol{Z\ test} = Z_{\bar{x}} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

Note, do not confuse the above with a Z score

$$\boldsymbol{Z\ score} = \frac{X - \bar{X}}{SD}$$

1. Find mean for sample of number of auto accidents in Statesboro during the past 10 years:

35, 41, 22, 51, 29,
16, 55, 61, 29, 43

M = 38.2

2. Find standard error of mean

Pop SD / SQRT(n)

Mean number of auto accidents in Statesboro during past 10 years = 38.2
Population Number of accidents during previous 20 years = 24.5
Population SD for accidents = 7.6
Sample size of 10 for previous ten years of data

$$Z_{\bar{x}} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \ ?$$

Pop SD = 7.6
N = 10

SE of mean = 7.6 / SQRT(10) = 2.403

[To obtain the square root when no square root function is available, raise value to .5, e.g., 7.6^.5]

3. Calculate Z score for sample vs population (or comparison) value

Pop M = 24.5
Sample M = 38.2
SE of Mean = 2.403

$$Z_{\overline{X}} = Z_M = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}}$$

Zm =

Zm = (38.2 – 24.5) / 2.403 =

Zm = 5.701

4. Determine whether this Zm is extreme – beyond what would be expected by random chance. How is this done?

One option is to calculate the p-value – probability of obtaining a sample mean that deviates this far, or further, from the population mean assuming Ho is true.

Use Z table to find the p-value for

Zm ≥ 5.70 or Zm ≤ -5.70

alternatively,

Zm ≥ |5.70|

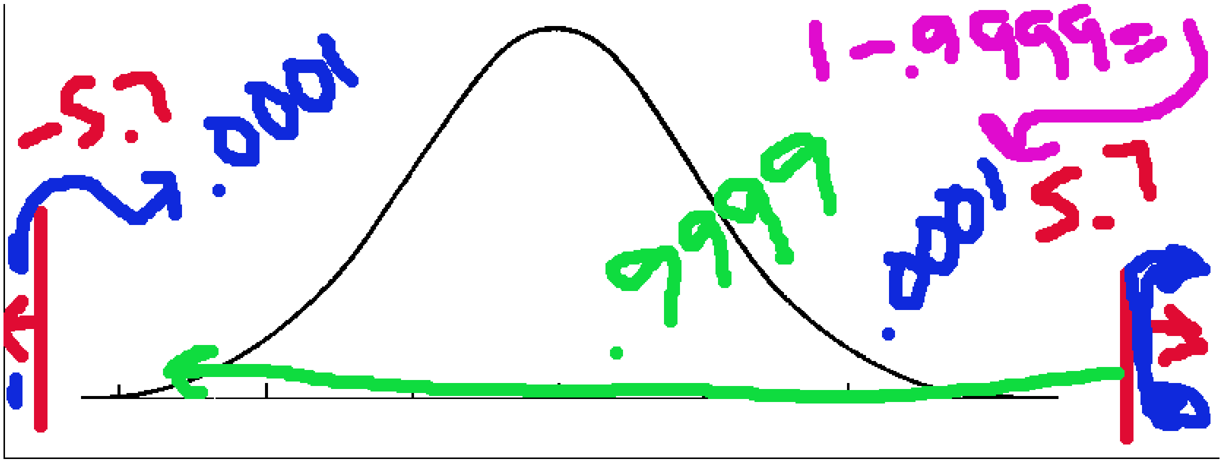
So,

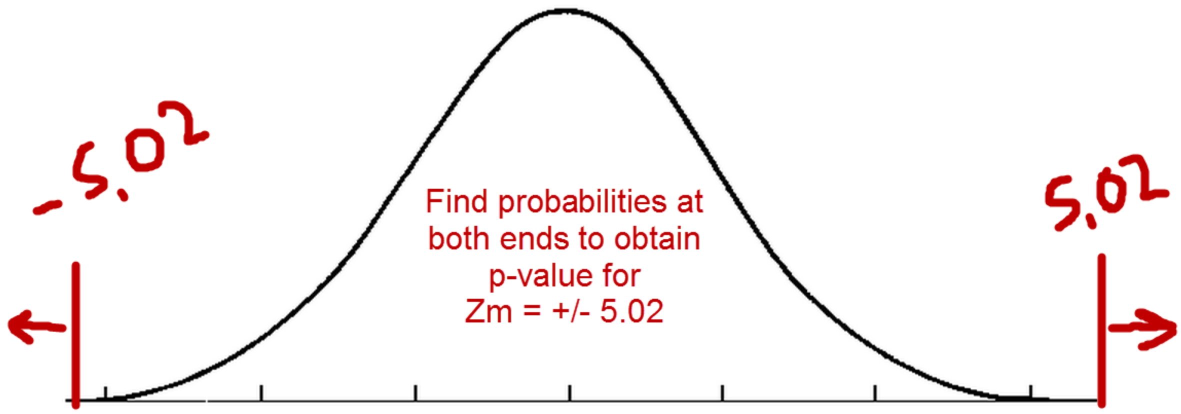
$$\text{p}(Z_{\bar{x}} \geq |calculated\ Z_{\bar{x}}|) = 2 \times probabilty\ from\ one\ tail\ of\ Z\ table$$

is the same as:

p(Zm ≥ 5.70 or Zm ≤ -5.70) =

p ≥ 5.7 = .0001
p ≤ -5.7 = .0001





p value for this test = .0002

How use this p value to determine whether Ho should be rejected?

Decision rule assuming α = .05

If p ≤ α reject Ho; if p > α fail to reject Ho
If p ≤ .05 reject Ho; if p > .05 fail to reject Ho

Decision rule more generally,

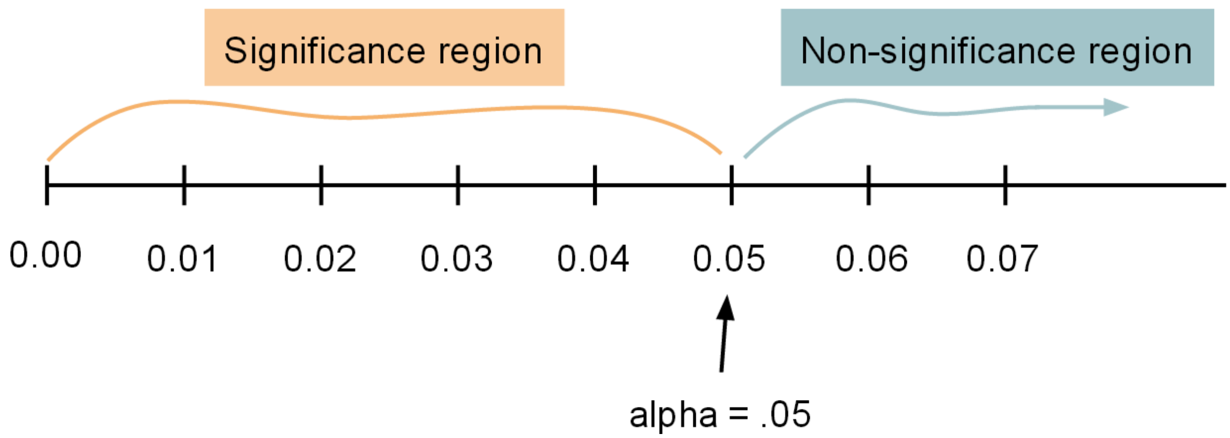
If p ≤ α reject Ho; if p > α fail to reject Ho

Plugging in the numbers:

If .0002 ≤ .05 reject Ho; if .0002 > .05 fail to reject Ho?

Reject Ho: μ = 24.5 and adopt H1 as descriptive of what was found.

Chart below shows how to interpret p-value when alpha = .05



Recall that **significant** means Ho was rejected

5. Inference and interpretation

Inference – tell reader whether Ho was rejected
Interpretation – tell reader what was found in simple language

Pop M = 24.5 (auto accidents previous 20 years)
Sample M = 38.2 (auto accidents past 10 years)
SE of Mean = 2.403
Zm = 5.7
p-value = .0002

Ho rejected? Use phrases like this:

"…found a statistically significant difference…"

Failed to reject Ho? Use phrasing like this:"

"...there was not a statistically significant difference…"
"...difference was not significant…"

Inference for current example:

"Analysis of the data revealed a statistically significant difference in mean number of auto accidents during the past 10 years when compared to number of accidents in the previous 20 years."

Remember, **significant means only that Ho was rejected**. It DOES NOT mean something important was found.

Example that Significant ≠ Important

Girl readings M = 85.3331313
Boys reading M = 85.3331312

This minor reading difference of 0.0000001 could be statistically significant if the sample size is large enough, but we all would agree this difference in reading scores is not important.

Interpretation for current one-sample Z test example:

"Data show that over the past 10 years the number of auto accidents in Statesboro appears to have increased from a mean number per year of 24.5 during the previous 20 years to a mean number of accidents of 38.2 for the past 10 years. "

Pop M = 24.5 (auto accidents previous 20 years)
Sample M = 38.2 (auto accidents past 10 years)

So the written result for the Z test would combine both the **Inference** and **Interpretation** statements, like this:

Analysis of the data revealed a statistically significant difference in mean number of auto accidents during the past 10 years when compared to number of accidents in the previous 20 years. Data show that over the past 10 years the number of auto accidents in Statesboro appears to have increased from a mean number per year of 24.5 during the previous 20 years to a mean number of accidents of 38.2 for the past 10 years.

In summary, for hypothesis testing using p-values, one compares p-value against alpha and if p is less than (or equal to) alpha, then reject Ho. This is true for any statistical test using p-values.

If p-value ≤ α reject Ho, but if p> α fail to reject Ho

Example 2 of Z-test

Reported ages

Student	Age
1	28
2	33
3	22
4	25
5	28
6	29
7	28
8	28
9	37

10	31
11	27
13	25

Mean Age for Students in Course = 28.4167

Population Age for GSU students = 24.5
Population SD for Age at GSU = 9.8
Sample size of 13 students from this course

$$Z_{\bar{x}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \quad ?$$

$$Z_{\bar{x}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{28.4167 - 24.5}{9.8/\sqrt{13}} = \frac{3.9167}{2.718} = \ 1.44$$

So what is the p-value for this Z-test score?

p($Z_{\bar{x}} \geq 1.44$) = 1 - .9251 = .0749
p($Z_{\bar{x}} \leq -1.44$) = .0749

Combined p = .0749+.0749 = .1498

p($Z_{\bar{x}} \geq |-1.68|$) $= 2 \times .0749 = \ .1498$

If α = .05, do we reject or fail to reject Ho?

Since .1498 is larger than .05, we fail to reject

If p-value ≤ α reject Ho, but if p> α fail to reject Ho
~~If .1498 ≤ .05 reject Ho~~, but if .1498> .05 fail to reject Ho

So what conclusion do we draw now about students in this course regarding their age relative to all students at GSU?

There is not a statistically significant mean difference between the sample mean and the population mean of 24.5. It appears that students in EDUR 8131 have an average age that is consist with GSU students overall.

3. One Sample Z-test with Critical Z

Critical Z values:

Alpha = .05
Z = **±1.96**

Alpha = .01
Z ≈ **±2.58**

Compare calculated Z against these critical Z values for hypothesis testing.

If $|Calculated\ Z| \geq |Critical\ Z|$ reject Ho otherwise fail to reject

Example:

Mean Age for Students in Course = 40.89
Ho: μ = 24.5

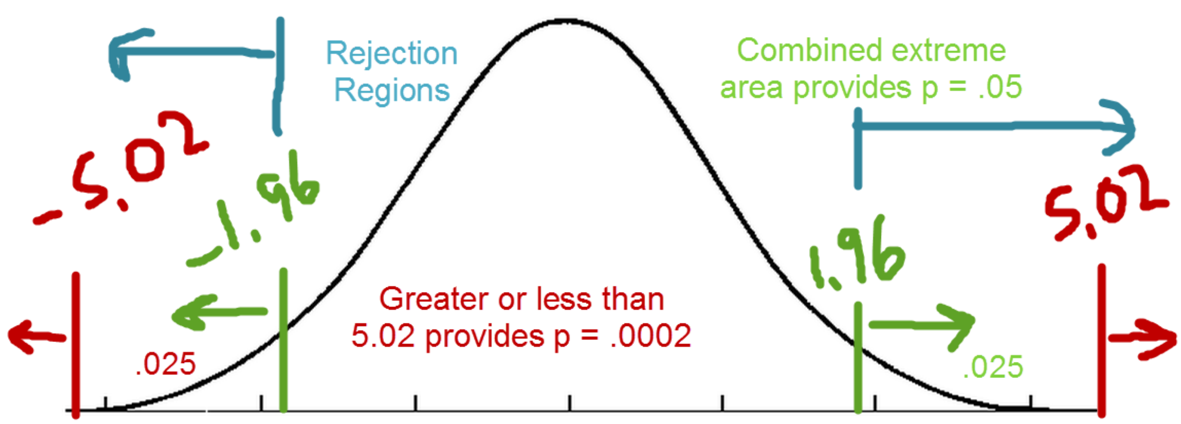
Population Age for GSU students = 24.5
Population SD for Age at GSU = 9.8
Sample size of 9 students from this course

$$Z_{\bar{x}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{40.89 - 24.5}{9.8/\sqrt{9}} = \frac{16.39}{3.267} = 5.02$$

Set alpha = .05, so critical Z scores are ±1.96

If $|Calculated\ Z| \geq |Critical\ Z|$ reject Ho otherwise fail to reject
If $|5.02| \geq |1.96|$ reject Ho otherwise fail to reject

Since 5.02 is larger than 1.96 we can reject Ho and draw the conclusion noted above.



Set alpha = .01, so critical Z scores are ±2.58

Would we reject or fail to reject Ho if the critical Z value is ±2.58?

If |5.02|≥ |2.58| reject Ho, otherwise fail to reject

So reject Ho since 5.02 is larger than critical value of 2.58

Example 2 of Z-test with Critical Values

Reported ages

Student	Age
1	28
2	33
3	22
4	25
5	28
6	29
7	28
8	28
9	37
10	31
11	27
13	25

Mean Age for Students in Course = 28.4167

Population Age for GSU students = 24.5
Population SD for Age at GSU = 9.8
Sample size of 13 students from this course

$$Z_{\bar{x}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \quad ?$$

$$Z_{\bar{x}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{28.4167 - 24.5}{9.8/\sqrt{13}} = \frac{3.9167}{2.718} = 1.44$$

If α = .01, this provides critical Z values of ±2.58. Do we reject or fail to reject Ho?

If |1.44|≥ |2.58| reject Ho, otherwise fail to reject
So fail to reject Ho since 1.44 is smaller than critical value of 2.58

If α = .05, this provides critical Z values of ±1.96. Do we reject or fail to reject Ho?

If $|1.44| \geq |1.96|$ reject Ho, otherwise fail to reject
So fail to reject Ho since 1.44 is smaller than critical value of 1.96

So what conclusion/interpretation do we draw now about students in this course regarding their age relative to all students at GSU?

Since Ho was not rejected, we can conclude that students in EDUR 8131 have an average age that is consistent with GSU students overall.

4. Assumptions of Z test

Normality and independence – see notes and video for presentation.

5. Errors in Hypothesis Testing

Type 1 error – we reject a **true** null (i.e., claim there is an effect based upon the sample selected when there is not an effect in the population); false positive

α = probability of a Type 1 error.

Type 2 error – failure to reject a **false** null (i.e., failing to find a difference or relationship in the sample when one actually exists in the population); false negative

β = probability of a Type 2 error.

Power – probability of rejecting a **false** null; power is the probability of finding an effect, finding a difference or relationship, if there is one in the population.

Power = 1 - **β**

Reading the table of hypothesis testing decisions

Population Situation Regarding H_0

One's Decision		H_0 True	H_0 False
	Reject H_0	Mistake (α) Type I error	Correct ($1 - \beta$)
	Fail to Reject H_0	Correct ($1 - \alpha$)	Mistake (β) Type II error

a. DV = math scores, IV = computer program usage. Experiment, half of 5th grade class uses computer program and other half does not. In the population of 5th grade students, the computer program results in a 20% increase in achievement based upon test scores. In your class you find evidence of achievement benefit in math scores due to the computer program so you reject Ho.

Null hypothesis states computer program has no benefit (i.e., math scores will not differ between those who use or do not use computer program).

Was this a correct decision or an error? If error, which type of error? Also, what is the probability of this decision?

Correct decision (probability is 1 – beta, which is called power)

Population Situation Regarding H_0

One's Decision		H_0 True	H_0 False
	Reject H_0	Mistake (α) Type I error	Correct ($1 - \beta$)
	Fail to Reject H_0	Correct ($1 - \alpha$)	Mistake (β) Type II error

b. DV = math scores, IV = computer program usage. Experiment, half of 5th grade class uses computer program and other half does not. In the population of 5th grade students, the computer program results in a 0% increase in achievement based upon test scores. In your class you find evidence of achievement benefit in math scores due to the computer program so you reject Ho.

Null hypothesis states computer program has no benefit (i.e., math scores will not differ between those who use or do not use computer program).

Was this a correct decision or an error? If error, which type of error? Also, what is the probability of this decision?

Type 1 error (probability is alpha)

Population Situation Regarding H_0

One's Decision		H_0 True	H_0 False
	Reject H_0	Mistake (α) Type I error	Correct ($1 - \beta$)
	Fail to Reject H_0	Correct ($1 - \alpha$)	Mistake (β) Type II error

c. DV = math scores, IV = computer program usage. Experiment, half of 5th grade class uses computer program and other half does not. In the population of 5th grade students, the computer program results in a 0% increase in achievement based upon test scores. In your class you find no evidence of achievement benefit in math scores due to computer program so you fail to reject Ho.

Null hypothesis states computer program has no benefit (i.e., math scores will not differ between those who use or do not use computer program).

Was this a correct decision or an error? If error, which type of error? Also, what is the probability of this decision?

Correct decision, and probability of this decision is 1-alpha

Population Situation Regarding H_0

One's Decision		H_0 True	H_0 False
	Reject H_0	Mistake (α) Type I error	Correct ($1 - \beta$)
	Fail to Reject H_0	Correct ($1 - \alpha$)	Mistake (β) Type II error

d. DV = math scores, IV = computer program usage. Experiment, half of 5th grade class uses computer program and other half does not. In the population of 5th grade students, the computer program results in a 10% increase in achievement based upon test scores. In your class, however, you find no evidence of achievement benefit in math scores due to computer program so you fail to reject Ho.

Null hypothesis states computer program has no benefit (i.e., math scores will not differ between those who use or do not use computer program).

Was this a correct decision or an error? If error, which type of error? Also, what is the probability of this decision?

Type 2 error (probability of this decision is beta)

The table below shows the hypothesis situations described above, but this table shows the situation in terms of symbolic hypotheses and numbers. Recall two groups compared, one with software and one without. Comparing mean achievement between the two groups.

Null – mean difference between groups is equal	Actual Mean Difference in Population	Researcher's decision based upon sample	Correct, Type 1, or Type 2 Error?	Probability of this outcome?
Ho: μ-μ= 0.00	100 – 100 = 0	Reject	Type 1	alpha
Ho: μ-μ= 0.00	120 – 100 = 20	Reject	Correct	1-beta
Ho: μ-μ= 0.00	100 – 100 = 0	Fail to Reject	Correct	1-alpha
Ho: μ-μ= 0.00	110 – 100 = 10	Fail to Reject	Type 2	beta

<u>Population Situation Regarding H_0</u>

<u>One's Decision</u>

	H_0 True	H_0 False
Reject H_0	Mistake (α) Type I error	Correct (1 - β)
Fail to Reject H_0	Correct (1 - α)	Mistake (β) Type II error

Power

See above; 1-β, probability of correctly rejecting a false null.