MATH 136 - Statistics Spring 2020

Exam 4

Ch 10, 11

Dr. Jorge Basilio

May_29 jbasilio@glendale.edu

Solutions

Directions

1. **At the top of your first page:** Please hand-write the statement provided below in quotes; print your name; put the date; and sign your name below it that acknowledges the **honor code**:

   "On my honor, by printing and signing my name, I vow to neither receive nor give any unauthorized assistance on this examination. I understand what my professor has deemed appropriate and inappropriate for this test and vow to follow these rules."

2. The exam is written to last 80 minutes, however, you have 3 hours to submit this exam without penalty.

3. The exam will be available on Canvas at 3 PM. You will need to submit your hand-written solutions by 6:00 PM.

4. How to submit: upload a **single PDF file** of your solutions to Canvas no later than 6:00 pm to avoid penalties.

5. Write your solutions to the exam on one side of the page (front side only, do NOT write double-sided). You do NOT need to copy the questions on your piece of paper. However, you must submit the test problems in the order given and you must clearly label each problem and part. If I cannot identify which problem you are working on, no points will be given.

6. Make sure to ALWAYS SHOW YOUR WORK; you will not receive any partial credits unless work is clearly shown unless told otherwise. *If in doubt, ask for clarification.* Correct answers with little to no work will receive no points. Students might be randomly selected to have a 1-1 conference where you are asked to defend your work and explain to me all your steps on certain questions or problems that are similar to a test question.

7. **Penalties for late submissions:** No late exams will be accepted. Exams received after 6:00 PM will not be graded and be given a score of 0.

8. **Allowed Materials**
   - You may use your calculator during the test (TI-83, 84, 84+, or 84+CET)
   - Blank pieces of paper to write your solutions. Writing utensils, erasers, etc
   - Formula Sheet (posted on Canvas and also at the end of this document)
   - Chi-Squared Table of Critical Values (posted on Canvas and also at the end of this document)

9. **Materials NOT Allowed**
   - Do not use your cell phone (for any reason: do not send or receive texts or calls or use the internet, etc)
   - Do not use your textbook (either digital or physical)
   - Do not use digital or printed out notes: the slides, the study guides, etc
   - Do not consult your HW
   - Do not give or receive any outside help (no getting help from a family member, friend, or any person either in person, via chat, message board, text message or any form of communication–again– you will be on camera the entire time so I will be looking for suspicious behavior)
   - Do not use your computer to look up anything using the internet (don't google; don't consult homework help websites, etc)

Continued on the next page...

Directions Continued:

10. The exam totals **135 points**.
11. There are 8 problems; many of them with multiple parts.
12. Handwriting should be neat and legible. If I cannot read your writing, zero points will be given.
13. Some questions contain multiple-parts which you must do individually and the parts are denoted by (a), (b), (c), etc. Some questions are multiple-choice and the choices are denoted with (A), (B), (C), (D), and (E). For True/False questions, you must spell out the entire word "true" or "false" in your answer.
14. Leave answers in exact form (as simplified as possible), unless told otherwise.
15. Put a $\boxed{\text{box around your final answer}}$ where applicable.
16. **PLEASE INCLUDE UNITS** where applicable
17. **PLEASE CHECK YOUR WORK!!!**
18. GOOD LUCK!!!!

Problem 1: 16 pts (2 pts each)

TRUE or **FALSE** (please spell out/write the entire word for credit). *(No work needed)*

(a) TRUE In a t-distribution when the degrees of freedom df increases, distributions becomes more like the z- distribution.

(b) TRUE The null hypothesis is a claim about a parameter assumed true until there is enough statistical evidence to reject it.

	H_0 T	H_0 F
FTR		II
R	I	

(c) FALSE A ~~type I~~ II error is made by failing to reject a false null hypothesis.

(d) TRUE All other things being equal, choosing a smaller value of α will increase the probability of making a type II error. β

(e) FALSE Two samples are said to be independent when the selection of the individuals in one sample ~~depends~~ doesn't depend on the selection of those in the other sample.

(f) FALSE If the null hypothesis is not rejected, there is ~~strong~~ statistical evidence that the null hypothesis is ~~proved~~ true. → no, we never "prove" the null

(g) TRUE A Type II Error is made when we fail to reject a false null hypothesis.

(h) TRUE We can use the calculator instructions **T-Test** to run a hypothesis test for one mean and also for two means from matched-pairs.

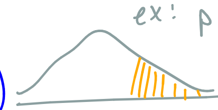
$\begin{cases} H_0: \mu = \mu_0 \\ H_A: \dots \end{cases}$ matched pairs $\begin{cases} H_0: \mu_d = 0 \\ H_A: \dots \end{cases}$

Problem 2: 24 pts (2 pts each blank)

Fill in the blanks:

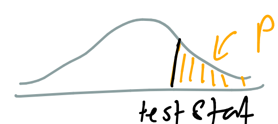
(a) In a right-tailed hypothesis test, the sign in the alternate hypothesis is greater than ($>$).

ex: $p > 0.5$



(b) If we get a p-value of 0.015 in a hypothesis test with a significance level of $\alpha = 0.02$, then we Reject the null hypothesis. so $p < \alpha$

(c) In a hypothesis test, the p-value is the probability of selecting a sample whose test statistic is at least as extreme as the observed test statistic that we got, assuming the null hypothesis is true.



(d) When converting two samples into one sample using the difference, d, of **matched-pairs**, we use the following notation:

Point Estimate: $\bar{d}$

True mean difference: μ_d Sample mean standard deviation: s_d

Null Hypothesis: $H_0: \mu_d = 0$ Alternate Hypothesis for a left-tailed test: $H_A: \mu_d < 0$

Requirements are: SRS and normally distributed or $n > 30$

Problem 3: 5 pts (2.5 pts each)

sorry formatting got screwed up :(

Multiple-choice. Select the correct answer. *(No work needed)*

(a) A confidence interval is an interval that is used to estimate a:

E

(A) sample statistic based on the information from a population.

(B) population parameter based on information from a population.

(C) sample statistic based on the information from a sample.

(D) sample parameter based on the information from a population statistic.

(E) population parameter based on information from a sample.

(b) When constructing a 95% confidence interval of the average (mean) yearly salary difference between females and males (independent) at a specific company (average female salary minus average male salary), our calculator gives us the interval $(-18240, -1115)$. This can be interpreted as

$\bar{x}_1 - \bar{x}_2$: avg female salary − avg. male salary $\bar{d} = \bar{x}_1 - \bar{x}_2 < 0$ bigger

(A) we are 95% confident that the females make between \$1,115 and \$18,240 ~~more than~~ males per year at this company.

B

(B) we are 95% confident that the females make between \$1,115 and \$18,240 less than males per year at this company.

(C) we are 95% confident that the females make between -\$1,115 and -\$18,240 per year at this company.

(D) we are 95% confident that the males make between \$1,115 and \$18,240 per year at this company.

(E) we are 95% confident that the males make between \$1,115 and \$18,240 per year at this company.

$\bar{x}_1 - \bar{x}_2 < 0$

$\bar{x}_1 < \bar{x}_2$

less

Problem 4: 10 pts (2 pts each)

Match the test with the scenario. *(No work needed)*
Write the letter of the hypothesis test or confidence interval from the list on the right that would be used for the situation on the left.

(a) E Cole wants to test (HT) if there is a difference (difference two μ, indep) in moisture content among turkeys that have been baked in the oven compared to those that have been deep fried. He takes two independent samples and measures their mean moisture level.

(b) J Nadine wants to estimate (CI) the average speed difference (difference two μ, indep) of cars on Highway 1 (between Santa Cruz and San Fransisco) and Highway 17. She selects two simple random, independent samples and measures their speeds.

(c) B Irene wants to test (HT) if the percentage (two proportions, indep) of students that own a smartphone is different than the percentage of faculty that own a smartphone at Glendale Community College.

(d) C Ethan wants to test (HT) if the average (mean) height of female volleyball players in the U.S. is more than 78 inches. (one μ)

(e) F Jesus wants to estimate (CI) the percentage of statistics students at Glendale Community College who took Math 30 or 30+. (one prop)

(A) Test about p

(B) Test about $p_1 - p_2$

(C) Test about μ

(D) Test about μ_d

(E) Test about $\mu_1 - \mu_2$

(F) Confidence Interval about p

(G) Confidence Interval about $p_1 - p_2$

(H) Confidence Interval about μ

(I) Confidence Interval about μ_d

(J) Confidence Interval about $\mu_1 - \mu_2$

(K) Confidence Interval about σ^2

(L) Confidence Interval about σ

Problem 5: 20 pts

→ sample $\hat{p} = 170/310$ → is coin fair? $p_0 = 0.5$

Students conducted an experiment to determine whether the Belgium-minted Euro coin was equally likely to land heads up or tails up. Coins were spun on a smooth surface, and in 310 spins, 170 landed with the heads side up.

→ making a claim H_1

Should the students interpret this result as convincing evidence that the proportion of the time the coin would land heads up is not 0.5? Test the relevant hypotheses using a significance level of $\alpha = 0.01$.

(3 pt) (a) Are the **requirements** met for the hypothesis test? Why or why not?

HT for one proportion

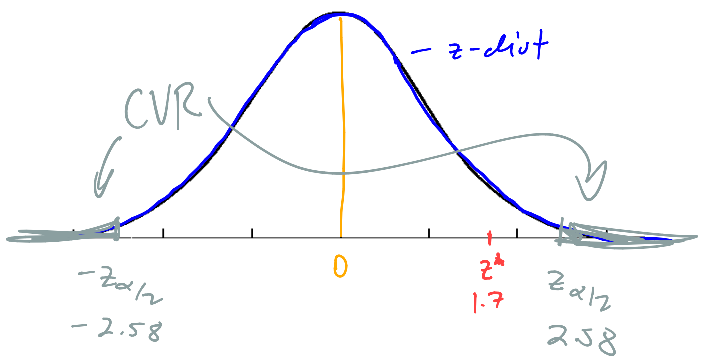
① SRS ② Requirements for binomial dist ③ $np_0 \geq 5$ $nq_0 \geq 5$ $310(0.5) \geq 5$ ✓

yes   i) fixed trials yes $n=310$ ii) independent yes   iii) two outcomes yes, H or T   iv) prob of success is constant. yes $p = 0.5$

(2 pt) (b) State the **null** and **alternative hypotheses**:

$$\boxed{\begin{cases} H_0: p = 0.5 \\ H_A: p \neq 0.5 \end{cases}}$$ (Two Tailed Test)

(1 pt) (c) State the **level of significance**: $\boxed{\alpha = 0.01}$



(3 pt) (d) Find the **test statistic** (show work!):

$$z^* = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.548 - 0.5}{\sqrt{\frac{(0.548)(0.452)}{310}}} = \boxed{1.70}$$

$$\hat{p} = \frac{170}{310} = 0.548 \quad \hat{q} = 1 - \hat{p} = 0.452$$

(4 pt) (e) Use the **Critical Value Method** to find the critical value:

Label the critical value and the shade the critical region in the distribution provided. Also, label the test statistic in your graph.

crit val $z_{\alpha/2}$

$\alpha = 0.01$

$\alpha/2 = 0.005$

$$z_{\alpha/2} = z_{0.005} = \text{invNorm}(0.005, 0, 1, \text{Right}) = \boxed{2.58}$$

(2 pt) (f) Make a **decision** (explain why):

Look at picture, we see $z^* = 1.70$ is outside the critical region (CVR)

Therefore, we $\boxed{\text{Fail to Reject } H_0}$

(5 pt) (g) $(M \to E)$ State your **conclusion**:

" There is not enough statistical evidence to support the claim that the proportion of heads of the Belgium-minted Euro coin is not 0.5. "

Problem 6: 20 pts

CI

same person : two means Matched-Pairs!

A researcher wanted to estimate the effect a new drug would have on systolic blood pressure. The following table gives the systolic blood pressure (in $mm\ Hg$) of seven adults before taking this drug, and after having taken this drug for 2 months.

		1	2	3	4	5	6	7
x_1	**Before**	210	180	195	220	231	199	224
x_2	**After**	195	178	186	223	218	195	224
$d = x_1 - x_2$	difference	15	2	9	-3	13	4	0

1-VAR STATS

$\bar{d} = 5.7$ mm Hg / person

$s_d = 6.8$ mm Hg

We will assume that the population of paired differences is normally distributed.

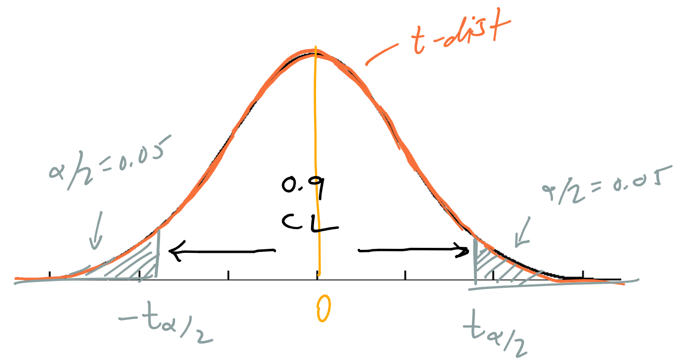
Construct a 90% confidence interval for the difference in systolic blood pressure before and after taking this drug.

mean

(4 pt) (a) Identify the **point estimate**:

parameter: mean difference of matched pairs

$\boxed{\bar{d} = 5.7 \text{ mmHg per person}}$



(4 pt) (b) Determine the **critical value**:

$CL = 0.9$

$\alpha = 1 - CL = 0.1$

$\alpha/2 = 0.05$

$t_{\alpha/2} = t_{0.05} = \left| \text{invT}(\underset{\alpha/2}{0.05}, \underset{df}{6}) \right|$

$= \boxed{1.94}$

(3 pt) (c) Find the **margin of error**:

$E = t_{\alpha/2} \cdot \dfrac{s_d}{\sqrt{n}} = (1.94) \cdot \dfrac{(6.8)}{\sqrt{7}} = \boxed{5.0 \text{ mmHg/person}}$

(4 pt) (d) Contsruct the **confidence interval**:

$CI: (\bar{d} - E, \bar{d} + E)$

$\bar{d} \pm E = 5.7 \pm 5.0$

$\boxed{CI: (0.7, 10.7)}$

(5 pt) (e) $(M \to E)$ **Interpretation of CI**:

"We are 90% confident that the difference in systolic blood pressure before & after taking this drug is between 0.7 mmHg and 10.7 mmHg per person."

Problem 7: 20 pts

MRT is MEAN

A Fair Isaac Corporation (FICO) score is used by credit agencies (such as mortgage companies and banks) to assess the creditworthiness of individuals. Values range from 300 to 850, with a FICO score over 700 considered to be a quality credit risk. According to Fair Isaac Corporation, the mean FICO score is 703. μ μ_0

A credit analyst wondered whether high-income individuals (incomes in excess of \$100,000 per year) had higher credit scores. He obtained a random sample of 40 high-income individuals and found the sample mean credit score to be 735 with a standard deviation of 80. s n $\bar{x}$

Test the claim that high-income individuals have higher FICO scores. HT ">"

(3 pt) (a) Are the **requirements** met for the hypothesis test? Why or why not?

HT: one mean

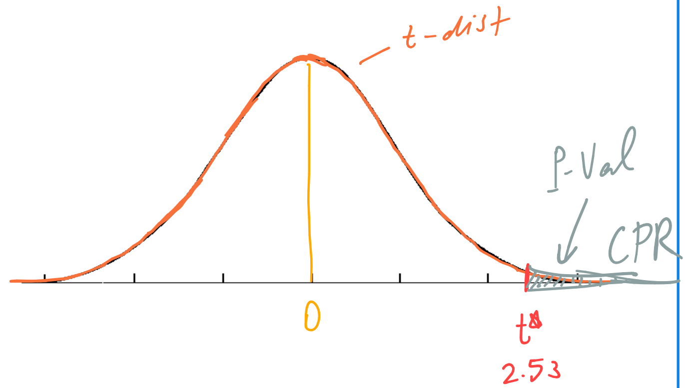
① SRS ✓ ② normally dist. or $n > 30$ ✓ $n=40$

Treat this as one mean b/c we aren't given two sample sizes.

(2 pt) (b) State the **null** and **alternative hypotheses**:

$$\begin{cases} H_0: \mu = 703 \ (\mu_0) \\ H_A: \mu > 703 \end{cases}$$

(Right Tailed Test)



(1 pt) (c) State the **level of significance**:

$\boxed{\alpha = 0.05}$ since it is not given

(3 pt) (d) Find the **test statistic** (show work):

$$t^* = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{735 - 703}{\frac{80}{\sqrt{40}}} = \boxed{2.53}$$

(4 pt) (e) Use the P**-Value Method** to find the P-value:
Label and shade the critical region in the distribution provided.

CPR: see picture

$$P(t > t^*) = P(t > 2.53) = \text{tcdf}(\underset{a}{2.53}, \underset{b}{1E99}, \underset{df}{39}) = \boxed{0.00778}$$

(2 pt) (f) Make a **decision** (explain why):

$\alpha = 0.05$

$P = 0.00778$

$p < \alpha \longrightarrow$ P low, null must go $\longrightarrow$ $\boxed{\text{Reject } H_0}$

(5 pt) (g) $(M \to E)$ State your **conclusion**:

"There is enough statistical evidence to support the claim that high income individuals have higher FICO scores."

Problem 8: 20 pts (5 each)

Short-response.
Instructions: For each question below, do the following parts:

(i) state the parameter;

(ii) determine if it is a Hypothesis Test or Confidence Interval (if it is a HT state the hypotheses; if it is a CI state the point estimate);

(iii) write the calculator instruction used;

(iv) use your calculator to answer the problem; either state the conclusion of a hypothesis test or give the confidence interval (no work necessary).

(a) A Pew Research Group conducted a poll in which they asked, "Are you in favor of, or opposed to, executing persons as a general policy when the crime was committed while under the age of 18?"

Of the 580 Catholics surveyed, 180 indicated they favored capital punishment; of the 600 seculars (those who do not associate with a religion) surveyed, 238 indicated they favored capital punishment.

claim HT

Is there a significant difference in the proportion of individuals in these groups in favor of capital punishment for persons under the age of 18?

$\neq$

i) parameter: difference of two proportions, independent

ii) HT or CI: HT : $\begin{cases} H_0: p_1 = p_2 \\ H_A: p_1 \neq p_2 \end{cases}$ (Two Tailed)

iii) Calc: 2PropZTest

iv) Answer: Calc $\Rightarrow$ $\begin{cases} z^* = -3.10 \\ p = 0.00194 \end{cases}$ $\alpha = 0.05$ (not given) $p < \alpha \rightarrow$ Reject H_0 $\rightarrow$ "There is enough evidence to support that there is a difference in the proportion of individuals in these groups in favor of capital punishments."

(b) Researchers wanted to determine if carpeted rooms contained more bacteria than uncarpeted rooms. To determine the amount of bacteria in a room, researchers pumped air from the room over a Petri dish at the rate of 1 cubic foot per minute for eight carpeted rooms and eight uncarpeted rooms. Colonies of bacteria were allowed to form in the 16 Petri dishes. The results are presented in the table. Assume that the data is approximately normally distributed.

claim HT

μ_1	Carpeted Rooms	11.8	10.8	7.1	14.6	8.2	10.1	13.0	14.0	L1
μ_2	Uncarpeted Rooms	12.1	12.0	3.8	10.1	8.3	11.1	7.2	13.7	L2

Do carpeted rooms have more bacteria than uncarpeted rooms at the $\alpha = 0.01$ level of significance?

μ_1: mean bacteria carpeted
μ_2: mean bacteria uncarpeted

i) parameter: two means, independent

ii) HT or CI: HT: $\begin{cases} H_0: \mu_1 = \mu_2 \\ H_A: \mu_1 > \mu_2 \end{cases}$ (Right-Tailed)

iii) Calc: 2-SampTTest

iv) Answer: $t^* = 0.96$ $\alpha = 0.01$ $p > \alpha \rightarrow$ P high, Null fly $\rightarrow$ Fail to Reject H_0

$p = 0.178$

$\rightarrow$ "There is not enough statistical evidence to support the claim that carpeted rooms have more bacteria than uncarpeted rooms."

(c) The Body mass index (BMI) of an individual is a measure used to judge whether an individual is overweight or not. A BMI between 20 and 25 indicates a normal weight.

In a random survey of 750 men and 750 women, the Gallup organization found that 203 men and 270 women were normal weight. Construct a 90% confidence interval to gaguge whether there is a difference in the proportion of men and women who are normal weight.

↳ CI ↳ independent

i) parameter: two proportions, independent

ii) HT or CI: CI : point estimate: $\hat{p}_1 - \hat{p}_2 = \frac{203}{750} - \frac{270}{750} = -0.089$ (men, women)

iii) Calc: $\boxed{\text{2PropZInt}}$

iv) Answer: $\boxed{(-0.129, -0.050)}$

$-0.129 < p_1 - p_2 < -0.050$

says: p_1 is smaller than p_2

so less proportion of men are normal weight.

(d) Octane is a measure of how much fuel can be compressed before it spontaneously ignites. Some people believe that higher-octane fuels result in better gas milage for their cars. To test this claim, a researcher randomly selected 11 individuals (and their cars) to participate in the study. Each participant received 10 gallons of gas and drove their car on a closed course that simulated both city and highway driving. The number of miles driven until the car ran out of gas was recorded. A coin flip was used to determine whether the car was filled up with 87-octane or 92-octane fuel first, and the driver did not know which type of fuel was in the tank.

claim HT

Matched Pairs

	Driver	1	2	3	4	5	6	7	8	9	10	11	
L1	Milage on 87 octane	234	257	243	215	114	287	315	229	192	204	547	→ x_1
L2	Milage on 92 octane	237	238	229	224	119	297	351	241	186	209	562	→ x_2

$d = x_1 - x_2$ (bigger)

claim: $d < 0$

i) parameter: two means, dependent (Matched Pairs)

ii) HT or CI: HT: $\begin{cases} H_0: \mu_d = 0 \\ H_A: \mu_d < 0 \end{cases}$

iii) Calc: $\boxed{\text{TTest}}$ using ↑ Need also $\bar{d} = -5.1$, $s_d = 14.9$

iv) Answer: $t^* = -1.14$ $\alpha = 0.05$ $p < \alpha \rightarrow$ P low, null go $\rightarrow$ $\boxed{\text{Reject } H_0}$

$p = 0.141$ (not given)

→ $\boxed{\text{"There is enough statistical evidence to support the claim that higher octane fuel results in better gas milage."}}$

Formula Sheet for Exam 4

- $\boxed{z = \dfrac{x-\mu}{\sigma}}$ $\boxed{x = \mu + z \cdot \sigma}$
- $\boxed{\begin{cases} \mu_{\bar{x}} = \mu \\ \sigma_{\bar{x}} = \sigma/\sqrt{n} \end{cases}}$
- $\boxed{\begin{cases} \mu_{\hat{p}} = p \\ \sigma_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n}} \end{cases}}$
- $\boxed{CL = 1-\alpha}$ $\boxed{\alpha = 1 - CL}$ $\alpha/2$

- $\boxed{\dfrac{(n-1)s^2}{\chi_R^2} < \sigma^2 < \dfrac{(n-1)s^2}{\chi_L^2}}$
- $\boxed{\chi_L^2 = \chi_{1-\alpha/2}^2}$ $\boxed{\chi_R^2 = \chi_{\alpha/2}^2}$
- $\boxed{\sqrt{\dfrac{(n-1)s^2}{\chi_R^2}} < \sigma < \sqrt{\dfrac{(n-1)s^2}{\chi_L^2}}}$
- $\boxed{\chi_L^2 = \chi_{1-\alpha/2}^2}$ $\boxed{\chi_R^2 = \chi_{\alpha/2}^2}$

Reminder: Consult the provided table for critical values for the χ^2-distribution

---

- $\boxed{\text{normalcdf}(a, b, \mu, \sigma)}$
- $\boxed{\text{invNorm}(\alpha, \mu, \sigma, \text{TAIL})}$
- $\boxed{z^* = \dfrac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}}$
- $\boxed{E = \mathbf{z}_{\alpha/2}\sqrt{\dfrac{\hat{p} \cdot \hat{q}}{n}}}$
- $\boxed{n = \dfrac{[z_{\alpha/2}]^2 \hat{p}\hat{q}}{E^2}}$
- $\boxed{n = \dfrac{[z_{\alpha/2}]^2}{4E^2}}$
- $\boxed{\text{tcdf}(a, b, \text{df})}$
- $\boxed{|\text{invT}(\alpha/2, \text{df})|}$
- $\boxed{t^* = \dfrac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}}$
- $\boxed{E = \mathbf{t}_{\alpha/2}\dfrac{s}{\sqrt{n}}}$
- $\boxed{n = \left[\dfrac{z_{\alpha/2} \cdot s}{E}\right]^2}$

- $\boxed{z^* = \dfrac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}}$
- $\boxed{\bar{p} = \dfrac{x_1 + x + 2}{n_1 + n_2}}$ $\boxed{\bar{q} = 1 - \bar{p}}$
- $\boxed{E = z_{\alpha/2}\sqrt{\dfrac{\hat{p}_1\hat{q}_1}{n} + \dfrac{\hat{p}_2\hat{q}_2}{n}}}$
- $\boxed{t^* = \dfrac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}}}$
- $\boxed{E = t_{\alpha/2}\sqrt{\dfrac{s_d}{n}}}$
- $\boxed{t^* = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}}$
- $\boxed{E = t_{\alpha/2}\sqrt{\dfrac{(s_1)^2}{n_1} + \dfrac{(s_2)^2}{n_2}}}$