Grade:______________

MIDTERM 2. MATH 362 SPRING 2014

SHOW YOUR WORK!. SHOW THE REASONING FOR YOUR SOLUTIONS AND IDENTIFY ALL VARIABLES TO RECEIVE THE FULL CREDIT! PLACE YOUR ANSWERS IN THE SPACE PROVIDED. SIMPLIFY COMPLETELY WHERE APPLICABLE.

---

(1) Consider independent random variables $Z \sim N(0,1)$, $U \sim \chi^2_n$ and $V \sim \chi^2_m$. Take $X_1, \dots X_k$ be a random sample from $N(\mu, \sigma^2)$ with sample standard deviation S_X.

Identify the distributions of the following quantities.

$$Z^2 \sim \underline{\chi^2_1}$$

$$\frac{Z}{\sqrt{U/n}} \sim \underline{t_n}$$

$$\frac{\bar{X}-\mu}{S_X/\sqrt{k}} \sim \underline{t_{k-1}}$$

$$\frac{k-1}{\sigma^2} S_X^2 \sim \underline{\chi^2_{k-1}}$$

$$\frac{V/m}{U/n} \sim \underline{F_{m,n}}$$

$$U+V \sim \underline{\chi^2_{m+n}}$$

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{k}} \sim \underline{N(0,1)}$$

(2) A university gave a validation test to entering students who had taken calculus in high school. The group of 13 students receiving no-college credit had a mean score of 4.17 on the test with a sample standard deviation of 3.70. For 7 students who received college credit from a high school dual enrollment program the mean score was 4.61 with a sample standard deviation of 2.28. For this problem let $\alpha = 0.05$.

(a) Identify the following quantities: (*Round to 2 decimal places for all calculations and show your calculations in empty space on the right*)

$\bar{X}$ = 4.17 $\qquad S_x^2$ = $(3.70)^2 = 13.69$

$\bar{Y}$ = 4.61 $\qquad S_y^2$ = $(2.28)^2 = 5.20$

n = 13 $\qquad S_p^2$ = 10.86

m = 7 $\qquad S_p$ = 3.30

(b) Test if the two samples have different variances by using a (two-sided) hypothesis test. Be sure to clearly state the test statistic, its distribution and your conclusion.

H_0 : $\sigma_X^2 = \sigma_Y^2$ versus H_1 : $\sigma_X^2 \neq \sigma_Y^2$

The test statistic is

$$F = \frac{S_Y^2}{S_X^2} = \frac{5.20}{13.69} = 0.38 \sim F_{6,12}$$

The critical values are

$$F_{0.025,6,12} = 0.186, \quad \text{and} \quad F_{0.975,6,12} = 3.73$$

Since $0.186 < 0.38 < 3.73$ we fail to reject H_0 and conclude that the variances of the two sets of student scores are equal.

Problem 2 continued.

(c) Test if there is a significant difference between the two groups by using a (two-sided) hypothesis test. Be sure to clearly state the test statistic, its distribution and your conclusion.

H_0 : $\mu_X = \mu_Y$ ______ versus H_1 : $\mu_X \neq \mu_Y$ ______

The test statistic is

$$T = \frac{\bar{X} - \bar{Y}}{Sp\sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{4.17 - 4.61}{3.3\sqrt{\frac{1}{13} + \frac{1}{7}}} = -0.28 \sim t_{18}$$

The critical values are

$$t_{0.025} = 2.10, \quad \text{and} \quad -t_{0.025} = -2.10$$

Since $-2.10 < -0.28 < 2.10$ we fail to reject H_0 and conclude that the means of the two sets of student scores are equal.

(d) Give the endpoints for a 95% confidence interval for the difference of the means. Be sure to state the formula you are using. Would you make the same conclusion?

The 95% CI for the difference of means is

$$\bar{X} - \bar{Y} \pm t_{0.025,18} \cdot S_p\sqrt{\frac{1}{n} + \frac{1}{m}} = (-3.69, 2.81)$$

Yes, we would also conclude that the means of the two sets of student scores are equal since the confidence interval contains zero.

(3) Let $X_1, \ldots, X_n$ be a random sample from a normal distribution with parameters μ and σ^2 where both μ and σ^2 are unknown.

(a) Derive the $(1-\alpha)100\%$ confidence interval for σ^2.

We know that

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$\begin{aligned}
1-\alpha &= P\left(\chi^2_{\alpha/2,n-1} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{1-\alpha/2,n-1}\right) \\
&= P\left(\frac{1}{\chi^2_{1-\alpha/2,n-1}} < \frac{\sigma^2}{(n-1)S^2} < \frac{1}{\chi^2_{\alpha/2,n-1}}\right) \\
&= P\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}\right)
\end{aligned}$$

Thus the $(1-\alpha)100\%$ confidence interval for σ^2 is given by

$$\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}, \ \frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}\right)$$

(b) Find the 95% confidence interval for σ^2 from data whose values gave $\bar{X} = 25$, $S^2 = 65.6, \ n = 16$.

$$\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}, \ \frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}\right) = \left(\frac{15(65.6)}{\chi^2_{0.975,15}}, \ \frac{15(65.6)}{\chi^2_{0.025,15}}\right) = (35.7969, 157.1348)$$

(4) Suppose that $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ are independent random samples from normal distributions with means μ_X and μ_Y and *known* variances σ_X^2 and σ_Y^2, respectively.

(a) Derive a $100(1-\alpha)\%$ confidence interval for $\mu_X - \mu_Y$.

By the CLT we know that

$$\bar{X} \sim N(\mu_X, \sigma_X^2/n) \quad \text{and} \quad \bar{Y} \sim N(\mu_Y, \sigma_Y^2/m)$$

we also know the additive/subtractive property of the normal distribution (can be proved by MGF) that

$$\bar{X} - \bar{Y} \sim N(\mu_X - \mu_Y, \sigma_X^2/n + \sigma_Y^2/m)$$

So,

$$\frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \sim N(0,1)$$

To derive the confidence interval, we consider

$$\begin{aligned}
1-\alpha &= P\left(-Z_{\alpha/2} < \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} < Z_{\alpha/2}\right) \\
&= P\left(-Z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} < \bar{X} - \bar{Y} - (\mu_x - \mu_y) < Z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right) \\
&= P\left(-(\bar{X} - \bar{Y}) - Z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} < -(\mu_x - \mu_y) < -(\bar{X} - \bar{Y}) + Z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right) \\
&= P\left(\bar{X} - \bar{Y} - Z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} < (\mu_x - \mu_y) < \bar{X} - \bar{Y} + Z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}\right)
\end{aligned}$$

Thus the $(1-\alpha)100\%$ confidence interval for $\mu_X - \mu_Y$ is given by

$$\bar{X} - \bar{Y} \pm Z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}$$

(b) Find the 95% confidence interval for $\mu_X - \mu_Y$ from data whose values gave $\bar{X} = 62, \bar{Y} = 52, n = 29,\ m = 20$ and known variances $\sigma_X^2 = 841$ and $\sigma_Y^2 = 400$.

$$\bar{X} - \bar{Y} \pm Z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} = 62 - 52 \pm Z_{0.05/2}\sqrt{\frac{841}{29} + \frac{400}{20}} = (-3.7197, 23.7198)$$

(c) From your confidence interval, would you conclude that the means are different? Explain.

No, the confidence interval for the difference of the means covers 0, so it is reasonable to believe that the means are equal.

(5) In a university study it was found that out of 50 math majors, 10 are left handed and out of 75 non-math majors only 5 were are handed. Does the 95% confidence interval for this study suggest that the proportion of left-handed math majors is different that that of non-math majors?

$$\begin{aligned} X &= \text{left-handed math majors} \\ Y &= \text{left-handed non-math majors} \end{aligned}$$

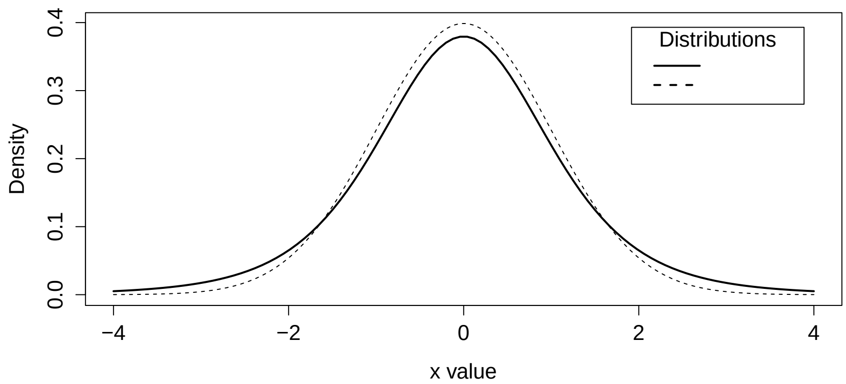
$$X = 10, \quad n = 50, \quad Y = 5, \quad m = 75$$

The 95% confidence interval is given by

$$\begin{aligned} &\frac{X}{n} - \frac{Y}{m} \pm Z_{0.025} \sqrt{\frac{\frac{X}{n}\left(1 - \frac{X}{n}\right)}{n} + \frac{\frac{Y}{m}\left(1 - \frac{Y}{m}\right)}{m}} \\ &= \frac{10}{50} - \frac{5}{75} \pm 1.96 \sqrt{\frac{\frac{10}{5}\left(1 - \frac{10}{50}\right)}{50} + \frac{\frac{5}{75}\left(1 - \frac{5}{75}\right)}{75}} \\ &= (0.0089, 0.2578) \end{aligned}$$

Yes, the 95% confidence interval suggest that there is a higher proportion of left-handed math majors because the interval is positive and does not contain the value 0.

(6) The graph below shows the standard normal distribution and a t-distribution with 5 degrees of freedom. Fill in the legend to correctly identify which curve is which. Explain how you chose your answer. Compare the critical regions for a hypothesis test using a t test statistic with one using a z test statistic. (*Are they the same? or is one bigger or smaller? and why?*)



The dashed line is the pdf of the standard normal distribution and the solid line is the t-distribution with 5 degrees of freedom. We know this because the tails of the t-distribution are thicker (due to the unknown variance) than that ones for the normal distribution.

You can also see this be comparing the rejection regions for a hypothesis test using the two distributions. Consider a two sided alternative hypothesis then the rejection region for the t-test we would have

$$|t| \geq t_{0.025,5} = 2.5706,$$

and for a z test statistic we would have

$$|z| \geq z_{0.025} = 1.96.$$

We can see that the fail to reject region for the t-distribution is a lot wider than that for the z-distribution. This a direct consequence of the fact that the variance is unknown and we have approximated it in the t-distribution leading to uncertainty in the width/spread of the data.