# STAT 303 Lecture 21 Outline: Formulas for Confidence Intervals and Tests of Hypotheses for One and Two Samples

H.J. Newton

March 4, 1996

## 1 Notation

- The symbol $Z_\alpha$ represents the value of a $Z$ random variable having area $\alpha$ to its right. You can get these numbers from the table on the inside front cover of Moore and McCabe, for example, $Z_{.05} = 1.645$ and $Z_{.10} = 1.29$.
- The symbol $t_{\alpha,v}$ represents the value of a $t$ random variable (having $v$ degrees of freedom) having area $\alpha$ to its right. These can be obtained from the table on the back inside cover of the text, for example, $t_{.10,7} = 1.415$.
- The symbol $\chi^2_{\alpha,v}$ represents the value of a $\chi^2$ random variable (having $v$ degrees of freedom) having area $\alpha$ to its right. These are in the table on page 816 of the text, for example, $\chi^2_{.10,7} = 12.017$.
- The symbol $F_{\alpha,v_1,v_2}$ represents the value of an $F$ random variable (having $v_1$ numerator degrees of freedom and $v_2$ denominator degrees of freedom) having area $\alpha$ to its right. These are in the table on page 810, for example, $F_{.05,7,9} = 3.29$.

## 2 Steps in Testing Hypotheses

1. Identify the parameter being discussed.
2. Formulate $H_0$ and $H_a$.
3. Specify $\alpha$, the risk you are willing to take of making a type I error.
4. Get data.
5. Calculate the value of the appropriate test statistic.
6. Reject $H_0$ if test statistic is too extreme (compare to a tabled value).
7. Make scientific conclusion.

## 3 Some Examples

We will go over the following problems in class.

1. In a study of the keyboard height that typists prefer, a sample of 31 trained typists was selected and their prefered heights determined. The resulting average height was 80 cm. Assuming that the population of all typists prefered heights is normally distributed with standard deviation 2 cm, test the hypothesis that the true mean prefered height is 79 cm.
2. A savings and loan association that finances a large number of home purchases in a particular region wanted information on the extent to which mortgage payments reduced the amount of disposable income during the initial year of occupancy for first-time home buyers. A sample of 35 recently granted mortgage applications was obtained, and for each the amount of the monthly mortgage payment as a percent of take-home income was computed. The sample average percentage was found to be 24.7, with a standard deviation of 3.2. Assuming that percentage is a normally distributed variable (among all applicants), test the hypothesis that the true average percentage is at least 22.
3. A recent newspaper article claims that at leaast 15% of all cars in California have had their emission control devices tampered with. A random sample of 200 cars were checked and it was found that 21 of them had in fact been tampered with. Does this justify the article's claim?
4. A ball bearing manufacturer claims he has developed a new process that leads to a reduction in the variation in the diameters in the ball bearings his plant produces. The current variance is 10. A random sample of 50 ball bearings is obtained using the new process and for these it is found that $s = 3$. Does this justify his claim?

5. A weight loss clinic claims its program reduces weight on the average by at least 10 pounds. A random sample of 8 patients have the following results:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Before | 160 | 170 | 235 | 145 | 263 | 167 | 205 | 184 |
| After | 148 | 163 | 215 | 138 | 240 | 155 | 197 | 172 |

Do these results justify the claim?

6. To compare two teaching methods, a school found 45 students who were as similar in backgrounds as they could and randomly selected 25 students to receive the first method and 20 students for the second. At the end of the school year, the same test was administered to both classes and the results recorded as follows: $\bar{X}_1 = 75$, $\bar{X}_2 = 79$, $s_1^2 = 9$, and $s_2^2 = 16$.

   (a) Assuming that the populations of test scores are normally distributed, test the hypothesis that the variability in the two populations is the same.

   (b) Based on your result in part (a), test the hypothesis that the two testing methods lead to the same average test score.

**Confidence Intervals and Tests of Hypotheses for One and Two Sample Inferences**

| Situation | | | |
|---|---|---|---|
| Confidence Interval | $H_0$ | Test Statistic | Rej $H_0$ if |
| For $\mu$, Normal sample, $\sigma^2$ known (`1 sample normal`) | | | |
| $\bar{X} \pm Z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$ | $\mu = \mu_0$ | $Z = \dfrac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ | $Z > Z_\alpha$<br>$Z < -Z_\alpha$<br>$\lvert Z\rvert > Z_{\alpha/2}$ |
| For $\mu$, Normal sample, $\sigma^2$ unknown (`1 sample t`) | | | |
| $\bar{X} \pm t_{\alpha/2,n-1}\dfrac{s}{\sqrt{n}}$ | $\mu = \mu_0$ | $t = \dfrac{\bar{X}-\mu_0}{s/\sqrt{n}}$ | $t > t_{\alpha,n-1}$<br>$t < -t_{\alpha,n-1}$<br>$\lvert t\rvert > t_{\alpha/2,n-1}$ |
| For $\mu_d$ = mean difference in paired data, Normal pairs, $\sigma_d^2$ unknown (`paired t`) | | | |
| $\bar{d} \pm t_{\alpha/2,n-1}\dfrac{s_d}{\sqrt{n}}$ | $\mu_d = \mu_0$ | $t = \dfrac{\bar{d}-\mu_d}{s_d/\sqrt{n}}$ | $t > t_{\alpha,n-1}$<br>$t < -t_{\alpha,n-1}$<br>$\lvert t\rvert > t_{\alpha/2,n-1}$ |
| For $\sigma^2$, Normal sample (`1 sample variance`) | | | |
| $\left(\dfrac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}, \dfrac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}\right)$ | $\sigma^2 = \sigma_0^2$ | $\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$ | $\chi^2 > \chi^2_{\alpha,n-1}$<br>$\chi^2 < \chi^2_{1-\alpha,n-1}$<br>$\chi^2 > \chi^2_{\alpha/2,n-1}$ or $\chi^2 < \chi^2_{1-\alpha/2,n-1}$ |
| For $\mu_1 - \mu_2$, $\sigma_1^2, \sigma_2^2$ known, independent normal samples (`2 sample normal`) | | | |
| $(\bar{X}_1 - \bar{X}_2) \pm Z_{\alpha/2}\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$ | $\mu_1 - \mu_2 = \delta$ | $Z = \dfrac{(\bar{X}_1 - \bar{X}_2) - \delta}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$ | $Z > Z_\alpha$<br>$Z < -Z_\alpha$<br>$\lvert Z\rvert > Z_{\alpha/2}$ |
| For $\mu_1 - \mu_2$, $\sigma_1^2, \sigma_2^2$ unknown but assumed equal, independent normal samples (`2 sample t`) | | | |
| $(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2,n_1+n_2-2}\sqrt{s_P^2(1/n_1 + 1/n_2)}$ | $\mu_1 - \mu_2 = \delta$ | $t = \dfrac{(\bar{X}_1 - \bar{X}_2) - \delta}{\sqrt{s_P^2(1/n_1 + 1/n_2)}}$ | $t > t_{\alpha,n_1+n_2-2}$<br>$t < -t_{\alpha,n_1+n_2-2}$<br>$\lvert t\rvert > t_{\alpha/2,n_1+n_2-2}$ |
| For $\pi$, sample from 0–1, $\min(n\pi, n(1-\pi)) \geq 5$, (`1 sample proportion`) | | | |
| $p \pm Z_{\alpha/2}\sqrt{p(1-p)/n}$ | $\pi = \pi_0$ | $Z = \dfrac{p - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}}$ | $Z > Z_\alpha$<br>$Z < -Z_\alpha$<br>$\lvert Z\rvert > Z_{\alpha/2}$ |

**Confidence Intervals and Tests of Hypotheses for One and Two Sample Inferences (cont)**

| Situation | | | |
|---|---|---|---|
| Confidence Interval | $H_0$ | Test Statistic | Rej $H_0$ if |
| For $\pi_1 - \pi_2$, independent samples from 0–1 populations, $\min(n_1\pi_1, n_1(1-\pi_1)) \geq 5$, $\min(n_2\pi_2, n_2(1-\pi_2)) \geq 5$, (`2 sample proportions`) | | | |
| $(p_1 - p_2) \pm$ | $\pi_1 - \pi_2 = \delta$ | $\dfrac{(p_1 - p_2) - \delta}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}}$ | $Z > Z_\alpha$ |
| $Z_{\alpha/2}\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}$ | $\pi_1 - \pi_2 = 0$ | $\dfrac{p_1 - p_2}{\sqrt{p(1-p)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$ | $Z < -Z_\alpha$ |
| | | $p = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$ | $\|Z\| > Z_{\alpha/2}$ |
| For ratio of variances, independent normal samples (`2 sample variance`) | | | |
| $\left(\dfrac{s_1^2/s_2^2}{F_{1-\alpha/2,n_1-1,n_2-1}}, \dfrac{s_1^2/s_2^2}{F_{\alpha/2,n_1-1,n_2-1}}\right)$ | $\sigma_1^2 = \sigma_2^2$ | $F = \dfrac{s_1^2}{s_2^2}$ | $F > F_{\alpha,n_1-1,n_2-1}$ |
| | | | $F < F_{1-\alpha,n_1-1,n_2-1}$ |
| | | | $F > F_{\alpha/2,n_1-1,n_2-1}$ or $F < F_{1-\alpha/2,n_1-1,n_2-1}$ |