

Stat312: Midterm Exam II.

Moo K. Chung, Yulin Zhang
mchung@stat.wisc.edu, yulin@stat.wisc.edu

November 11, 2004

Please use the following R output in answering the appropriate exam problems.

```
> pnorm(c(1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6))
[1] 0.84 0.86 0.88 0.90 0.92 0.93 0.95
> pt(c(1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6), 9)
[1] 0.83 0.85 0.87 0.89 0.90 0.92 0.93
> pt(c(1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6), 10)
[1] 0.83 0.85 0.87 0.89 0.90 0.92 0.93
```

1. A mixture of pulverized fuel ash and Portland cement should have a compressive strength of more than 1000KN/m³. The mixture will not be used unless experimental evidence indicates that the strength specification has been met. Suppose compressive strength for specimens of this mixture is normally distributed with $\sigma^2 = 98^2$. The sample mean of compressive strength for 10 specimens is $\bar{x} = 1050$.

(1) What are the appropriate null and alternate hypothesis (5 points)?

(2) What would be the appropriate test statistic for the hypothesis testing? What is the probability distribution of your test statistic when H_0 is true? (5 points each)

(3) Do hypothesis testing at 85% level. Has the strength specification been met (10 points)?

(4) Compute the P -value (5 points).

Solution. (1) Let μ be the population mean of compressive strength. Then $H_0 : \mu = 1000$ and $H_1 : \mu > 1000$. (2) Since the population variance is known, the test statistic should be the Z statistic $Z = (\bar{X} - \mu)/(\sigma/\sqrt{n})$. Under H_0 , $Z \sim N(0, 1)$. (3) $z = 50/(98/\sqrt{10}) = 1.61$. From R output, 85% level rejection region is $z > 1.1$. So we reject H_0 . (4) $P\text{-value} = P(Z > 1.61) = 1 - P(Z \leq 1.61) = 0.05$.

2. A farmer want to know if there is any difference

in weight gain from young male and female pigs when both sexes are subjected to identical feeding program. He measured the weight gain for 7 female pigs (x_i) and 5 male pigs (y_i) and observed $\bar{x} = 9.8, \bar{y} = 9.2, s_x = 0.62, s_y = 0.86$. s_x and s_y are the sample variance of female pigs and male pigs respectively.

(1) State the appropriate null and alternate hypothesis (5 points).

(2) Compute the pooled sample variance (5 points).

(3) What is the probability distribution of the test statistic under H_0 ? What are the statistical assumptions on measurements to get this distribution? (5 points each).

(4) Compute the P -value. Based on the P -value, do hypothesis testing at $\alpha = 0.05$ (5 points each).

Solution. (1) Let μ_X and μ_Y be the weight gains for female and male pigs respectively. Then $H_0 : \mu_X = \mu_Y$ vs. $H_1 : \mu_X \neq \mu_Y$. (2) The pooled variance is

$$s_p^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2} = 0.53.$$

(3)

$$T = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{1/n + 1/m}} \sim t_{10}.$$

We need to assume that x_i come from $N(\mu_X, \sigma^2)$ and y_j come from $N(\mu_Y, \sigma^2)$. Further, x_i and y_j should be independent observations. (4) $P\text{-value} = P(|T| > 1.41) = 2(1 - P(1.41)) = 2(1 - 0.9) = 0.2$. Since our P -value is bigger than the given α , we do not reject H_0 .

3. A study found that of 238 individuals with coronary heart disease (CHD), 145 smoked cigarettes; of 476 in the same age group but **without** CHD, 192 smoked cigarettes. The researchers want to know if the proportion of cigarettes smokers is the

same in the population of individuals with CHD as in the population with no CHD.

- (1) State the null and alternate hypothesis. (5 points)
- (2) What's your test statistic and its distribution under the null hypothesis? (5 points)
- (3) Perform the hypothesis test using a significance level $\alpha = 0.14$. (10 points)
- (4) Calculate a 86% confidence interval for the difference between the two proportions. (10 points)

Solution. (1) Let p_C and p_N be the proportions of smokers in CHD group and non CHD group respectively. Then $H_0 : p_C = p_N$ vs. $H_1 : p_C \neq p_N$. (2) The point estimator for $p_C - p_N$ is $\hat{p}_C - \hat{p}_N = X/n - Y/m$, where X, Y are the total number of smokers in CHD group and non CHD group respectively and $n = 238, m = 476$. Then under $H_0 : p_C = p_N = p$, the test statistic is

$$Z = \frac{X/n - Y/m}{\sqrt{p(1-p)(1/n + 1/m)}} \sim N(0, 1).$$

Since p is unknown, you estimate it by pooling the samples: $\hat{p} = (x + y)/(m + n)$. (3) $\hat{p} = (145 + 192)/(238 + 476) = 0.472$ and $z = 5.20$. We reject H_0 if $|z|$ is large. Since $\alpha = 0.14$, the cutoff value should be $z_{\alpha/2} = z_{0.07} = 1.5$ from the R output. (4) Although `qnorm` values are not given, we can estimate $z_{0.07}$ from the R output. Note that $P(Z > z_{0.07}) = 0.07$. So $z_{0.07} = 1.5$. A $100(1 - \alpha)\%$ confidence interval is

$$\hat{p}_C - \hat{p}_N \pm z_{\alpha/2} \sqrt{\hat{p}_C(1 - \hat{p}_C)/238 + \hat{p}_N(1 - \hat{p}_N)/476}.$$

With $\hat{p}_C = 145/238$ and $\hat{p}_N = 192/476$, we get $[0.15, 0.26]$.