Welcome to your last homework devoted to practical applications of the theory of hypothesis testing. Exam 3 is next Thursday.

As usual, try to find mistakes (and get extra points) in my solutions. Typically they are silly arithmetic mistakes (not methodological ones). They allow me to check that you did your HW on your own. Please do not e-mail me about your findings — just mention them on the first page of your solution and count extra points.

Now let us look at your problems.

1. Problem 13.20. If p-value (observed level of significance) is .0316 (3.16%) then:
   a. at $\alpha = .01$, $H_0$ is not rejected.
   b. at $\alpha = .05$ the null hypothesis is rejected,
   c. at the .1 level of significance the null hypothesis is rejected.

   Recall the general rule: p-value should be relatively small (with respect to the level of significance) to reject the null hypothesis.

2. Problem 13.24. Given: $\alpha = .05$, $n = 16$, $\mu_0 = 10$, $\bar{x} = 8.4$, $\sigma = 3.2$, $H_0 : \mu = \mu_0 = 10$ versus $H_a : \mu < \mu_0$.

   Here the test statistics is
   $$ z = \frac{\bar{x} - \mu_0}{\sigma/n^{1/2}} = \frac{8.4 - 10}{3.2/4} = \frac{-1.6}{3.2} = -2. $$

   Because $z_{.95} = -1.645$, the answer is “Reject $H_0$”. Also,

   $$ p-value = P(Z > |z|) = P(Z > 2) = .5000 - .4772 = .0228. $$

   Because the p-value is relatively small (smaller than the level of significance), we confirm our “Reject” decision.

3. Problem 13.34. Among 10,000 people there are $n_c$ with cancer and $(10,000 - n_c)$ without. For each participant in that study an epidemiologist observes 48 factors (observations) $F_i, i = 1, 2, \ldots, 48$. [Note that the epidemiologist has a matrix 10,000 times 49 of data with each row containing in the first column 1 or 0 (cancer or no cancer) and in others values of 48 factors.

   Then, using this matrix of data, she calculates for each factor (column) an average over people with cancer $\bar{F}_{ic}$ and without cancer (healthy) $\bar{F}_{ih}$. Please note that here we have the case of two populations: with and without cancer. Then for each factor we consider a classical hypothesis testing problem about a factor’s population mean: $H_{a0} : \mu_{ic} = \mu_{ih}$ versus $H_{ia} : \mu_{ic} \neq \mu_{ih}$. Note that the null hypothesis means that an $i$th factor does not cause cancer [say, if $F_i$ is weight then people with cancer and no-cancer have the same population weight and then this is not a causal factor in cancer]. Further, she considers all 48 two-populations tests with the level of significance .01. (Do you recall a Bonferroni rule that should be used in such a case of simultaneous hypotheses testing? Here the epidemiologist forgot about this
but you, the statisticians, may help her to correctly solve the problem of choosing the level of significance for each test.)

A conclusion that \( i \)th factor is associated with the cancer, even if none of them is associated (is a causal factor) means that \( H_0 \) is rejected (sample factor’s means are significantly different) while the null hypothesis for \( i \)th factor is true (population means are the same). Note that this error occurs with probability \( \alpha = .01 \).

As a result, let \( X \) be a number of such “associated with the cancer” causal factors caused by a stochastic nature of the data. Note that under the null hypotheses \( X \sim \text{Binom}(\alpha, 48) \). As a result,

\[
P(X = 1) = \frac{48!}{(1!)(47!)} \alpha^1 (1 - \alpha)^{47} = 48\alpha(1 - \alpha)^{47}.
\]

(b) Here

\[
P(X > 1) = 1 - P(X = 0) - P(X = 1) = 1 - (1 - \alpha)^{48} - 48\alpha(1 - \alpha)^{47}.
\]

4. Problem 1.36. Here we have two populations: insurance and banking industries. Given: \( n_1 = 40, \bar{x}_1 = 9.1, s_1 = 1.9, n_2 = 50, \bar{x}_2 = 8, s_2 = 2.1 \). We test \( H_0 : \mu_1 - \mu_2 = 0 \) versus \( H_a : \mu_1 - \mu_2 \neq 0, \alpha = .05 \). Please note that sample sizes are large enough. Consider test-statistic

\[
z = \frac{\bar{x}_1 - \bar{x}_2}{[s^2_1/n_1 + s^2_2/n_2]^{1/2}} = \frac{9.1 - 8}{[1.9^2/40 + 2.1^2/50]^{1/2}} = \frac{1.1}{1.1/422} = 2.61.
\]

From the Normal Table, \( z_{\alpha/2} = z_{.025} = 1.96 \). This implies “Reject the null hypothesis". Further,

\[
p - value = 2P(Z > 2.61) = 2(.0045) = .009.
\]

This small p-value confirms our rejection decision. You need to request something like .5% level of significance to get “Do not Reject“ decision.

5. Problem 13.40. This is again the case of two populations, but here samples are small so the authors are forced to assume that the distributions are normal with the same variances. Given: \( \bar{x}_1 = 77.4, s_1 = 3.3, n_1 = 6, \bar{x}_2 = 72.2, s_2 = 2.1, n_2 = 6, \alpha = .01, H_0 : \mu_1 = \mu_2 \) versus \( H_a : \mu_1 \neq \mu_2 \).

The pooled sample variance is

\[
s^2_p = \frac{(n_1 - 1)s^2_1 + (n_2 - 1)s^2_2}{n_1 + n_2 - 2} = \frac{5(3.3^2 + 2.1^2)}{10}
\]

\[
= \frac{10.89 + 4.41}{2} = 15.3/2 = 7.65.
\]

This yields \( s_p = 2.77 \) (please note that it is between \( s_1 \) and \( s_2 \)). Then

\[
t = \frac{\bar{x}_1 - \bar{x}_2}{s_p[1/n_1 + 1/n_2]} = \frac{77.4 - 72.2}{(2.77)(.577)} = \frac{5.2}{4.41} = 3.25.
\]
From the T-Table we get $t_{\alpha/2,10} = 2.228$. Thus we reject the null hypothesis.

6. Problem 13.48. Here $n_1 = 24$, $s = 238$. Assuming a normal population, test $H_0 : \sigma = \sigma_0 = 250$ versus $H_a : \sigma \neq 250$ with $\alpha = .01$. Here the test statistic is

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(23)(238)^2}{250^2} = \frac{(23)(56644)}{62500} = 20.845.$$  

Then we reject if either $\chi^2 < \chi^2_{1-\alpha/2,10}$ or $\chi^2 > \chi^2_{\alpha/2,10}$. Here $s < \sigma_0$ so we are looking at the left tail. From the chi-squared Table $\chi^2_{1-\alpha/2,10} = \chi^2_{.995,25} = 9.260$. The conclusion is do not reject.

7. Problem 13.54. This is the case of two populations where we are interested in a conjecture about population variances, namely $H_0 : \sigma_1 = \sigma_2$ versus $H_a : \sigma_1 \neq \sigma_2$. Here $s_1 = 3.3$, $n_1 = 6$, $s_2 = 2.1$, $n_2 = 6$, $\alpha = .1$. The test statistic is

$$f = \frac{s_1^2}{s_2^2} = \frac{3.3^2}{2.1^2} = 2.47.$$  

From the Table $f_{\alpha/2,5.5} = 5.05$. We conclude that there is no evidence for rejecting the null hypothesis that population sample variances are the same. Note: this test justifies the assumption used in 13.40 that the variances are the same.

8. Problem 13.57. Here $x = 10$, $n = 18$, $H_0 : \theta = .4$ versus $H_a : \theta > .4$. Test with $\alpha = .05$. It is convenient to do this via p-value (I will use the Binom. Table):

$$p-value = P(X_b \geq 10|\theta = .4, n = 18) = .0771 + .0374 + \ldots$$  

I can stop here because obviously the p-value is larger $\alpha$ so the answer is “Do not Reject”. By the way, using the normal approximation, which is OK here, we get

$$P(X_b \geq 10) = P\left(\frac{X_N - n\theta_0}{\sqrt{n\theta_0(1 - \theta_0)}} \geq \frac{9.5 - n\theta_0}{\sqrt{n\theta_0(1 - \theta_0)}}\right)$$

$$= P(Z \geq 2.3/2.078) = P(Z \geq 1.11) = 1.225$$  

It is close to the answer in your text based on the Binomial Table, I guess.

9. Problem 13.59. Here $x = 1$, $n = 19$, $\theta_0 = .3$ with $H_0 : \theta = \theta_0$ versus $H_a : \theta < .3$, and $\alpha = .05$. Using the Binomial Table,

$$p-value = P(X_b \leq 1|\theta = .3, n = 19) = .0011 + .0093 = .0104.$$  

As a result, the observed level of significance is smaller than the level of significance, and the null hypothesis should be rejected.

Question: Suggest $\alpha$ that the answer is “Do not reject”.

10. Problem 13.63. Here $n = 1000$, $x = 290$, $H_0 : \theta = .35$ $H_a : \theta < .35$, $\alpha = .05$. Let us look at p-value:

$$p-value = P(X_b \leq 290) = P(X_N \leq 290.5)$$
\[ P(\frac{X_N - n\theta}{[n\theta(1 - \theta)]^{1/2}} \leq \frac{290.5 - 350}{[(350)(.65)]^{1/2}}) = P(Z < -59.5/15.08) = P(Z > 3.95) = 0.\]

Thus we reject the null hypothesis for any level of significance.

11. Problem 13.73. Here we have 4 Binomial trials with \(n_1 = n_2 = n_3 = n_4 = 200\), and observations (numbers of successes or, in the “table literature” terminology, frequencies) \(f_{11} = 26, f_{21} = 23, f_{31} = 15\) and \(f_{41} = 32\). The null hypothesis is that the probabilities of successes are the same while the alternative that this is not the case. Here a chi-squared test is absolutely appropriate. Calculations are not “nice and easy” but manageable.

Our first step is to calculate, under the null hypothesis, the probability of success (note that we pool together all observations):

\[
\hat{\theta} = \frac{f_{11} + f_{21} + f_{31} + f_{41}}{n_1 + n_2 + n_3 + n_4} = \frac{26 + 23 + 15 + 32}{(4)(200)} = .12.
\]

Then, according to section 13.6, we calculate chi-squared statistic

\[
\chi^2 = \sum_{i=1}^{3} \sum_{k=1}^{2} \frac{(f_{ik} - e_{ik})^2}{e_{ik}}.
\]

Here \(f_{i2} = n_i - f_{i1}, e_{i1} = n_i\hat{\theta}\) and \(e_{i2} = n_i(1 - \hat{\theta})\). We plug-in and get

\[
\chi^2 = \frac{(26 - 24)^2}{24} + \frac{(174 - 176)^2}{176} + \frac{(23 - 24)^2}{24} + \frac{(177 - 176)^2}{176} + \frac{(15 - 24)^2}{24} + \frac{(185 - 176)^2}{176} + \frac{(32 - 24)^2}{24} + \frac{(168 - 176)^2}{176}
\]

\[
= 4 + 1 + 81 + 64 + 4 + 1 + 81 + 64 = \frac{200}{24} = 7.10.
\]

From the Table we get \(\chi^2_{05,3} = 7.815\), so we do not reject the null hypothesis.