Solutions to Homework #9, 36-220

Due 16 November 2005

8.54

We are supposed to test whether "the true average percentage of organic matter in such soil is something other than 3%". So our null hypothesis H_0 is that $\mu = 3$. We need a two-sided test, since we want to tell whether the true percentage is *different* than 3%, whether the difference is positive or negative. So H_1 is that $\mu \neq 3$. Because σ is not known, and n is comparatively small, only 30, we can't use a z (standard normal) test. We can however use a t test, because we're given that the individual measurements are normal.

$$t = \frac{2.481 - 3}{0.295} = -1.76$$

Since $n = 30$, there are $30 - 1 = 29$ degrees of freedom. To compute the p-value, we need the probability that a t-distribution with 29 degrees of freedom would give a value as far from the origin as we got — $\Pr(|T| \geq |t|)$. This is

$$p = \Pr(T \leq -|t|) + \Pr(T \geq |t|) = 2\Pr(T \geq |t|)$$

since the t-distribution is symmetric. Turning to appendix A.8, we see that, with 29 degrees of freedom, $\Pr(T \geq 1.7) = 0.05$, and $\Pr(T \geq 1.8) = 0.041$. Linear extrapolation between these values gives $\Pr(T \geq 1.76) = 0.0446$. (Using statistical software gives 0.04456319.) That is, according to the null hypothesis, we should expect to see results this far out along *one* of the tails 4.46 percent of the time. Since we're using a two-sided test, we need to double this probability, $p = 0.0892$. This is less than 0.1, so we'd reject the null at the 10% level. At that level, we do have evidence that the mean concentration of organic matter in the soil isn't 3%. At the stricter $\alpha = 0.05$ level, we don't have enough evidence to reject the null that $\mu = 3$.

8.58

The SAS includes a two-sided p-value, namely 0.0711. We want to perform a two-sided test — there will be a problem with our ciruits if the cut-on voltage of the diodes is either too high or too low — so this suits us just fine. Because

$0.0711 > 0.01$ and > 0.05, we conclude that, *at that significance level*, cut-on voltages do not differ systematically from 0.60V; with those values of α, we do not need to adjust anything. But $0.0711 < 0.1$, so the deviation from the null is significant at the less reliable 10% level. If that is our chosen value of α, we must adjust, because the average cut-on voltage is too high (the calculated test statistic t is positive).

8.61

We are supposed to test whether the average thickness of the lenses "is something other than what is desired"; deviations in both directions are bad, so we need a two-sided test. The null is H_0: $\mu = 3.20$mm, the desired width. The alternative then is H_1: $\mu \neq 3.20$mm. We do not know that the individual lens-widths are Gaussian, or what the population standard devation is, but because n is large enough (50), we should be safe using a z-test anyway.

$$z = \frac{3.05 - 3.20}{0.34/\sqrt{50}} = -3.12$$

This gives a p-value of 0.00181. This is well below $\alpha = 0.05$, so we have good reason to think that the average lens width is different from the desired, and in fact smaller.

Just as a check as to whether or not using the z-test was OK, let's calculate the p-value with a t-test. We shouldn't *really* use a t-test here, since to do so we'd have to know that the individual observations are Gaussian. But if the t-test and the z-test give approximately the same p-values, that's a good indication that we've got enough measurements that the *sample mean* is Gaussian, which is all we really care about. So, plugging the data into the t-test, we get $t = -3.12$, just as above, with $50 - 1 = 49$ degrees of freedom. The p-value is then 0.003, which is close enough to what we got with the z-test for all practical purposes.

8.62

We think that we know $\sigma = 0.30$. Therefore, in computing β, we can use the z-test formulas. (Since the observed s is 0.34, which is pretty close to our hypothesized σ, we seem to know what we're doing.) Those formulas are given on page 330.

$$n = \left[\frac{(z_{\alpha/2} + z_\beta)\sigma}{\mu_0 - \mu}\right]^2$$

In our case, $\alpha = 0.05$ so $z_{\alpha/2} = z_{0.025} = 1.96$, $\beta = 0.05$ and $z_\beta = 1.64$, $\mu_0 = 3.20$, and $\mu = 3$. Thus, $n = 29.2$, or 30, rounding up. So the actual sample size (50) was two-thirds larger than it needed to be, in order to get this much power to detect the alternative $\mu = 3.0$.

9.10

Let Δ be the amount by which the average toughness of high-purity steel exceeds the average toughness of commercial steel. The null hypothesis is that $\Delta = \Delta_0 = 5$. The relative alternative is that $\Delta > 5$. Because we assume both populations are Gaussian, and that we know σ_1 and σ_2, we can use the z test, regardless of the population size.

a

The test statistic is

$$z = \frac{\overline{x_1} - \overline{x_2} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{65.6 - 59.8 - 5}{\sqrt{\frac{(1.2)^2}{32} + \frac{(1.1)^2}{38}}} = 2.89$$

The p-value is 0.0019, which is greater than 0.001, so we cannot reject the null hypothesis at that level. In terms of the original problem, while it's not very likely, under the null hypothesis, that we'd see this big a gap between the two types of steel, it's not *so* unlikely that it meets the standard of confidence we've decided we need.

b

The formula for the "miss" rate, with a one-sided test, is (p. 365)

$$\beta(\Delta) = \Phi\left(z_\alpha - \frac{\Delta - \Delta_0}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}}\right)$$

$\alpha = 0.001$, so $z_\alpha = 3.09$. $\Delta_0 = 5$, and, in this particular alternative, $\Delta = 6$. $\sigma = 0.28$. Putting this all together, we get $\beta(6) = \Phi(-0.53) = 0.29$. So this test has a pretty high probability of missing this deviation from the null, even if it's really happening.

9.40

To do a paired test, we need the mean and standard deviation of the differences. For each period i, calculate d_i as the number of fish caught with the pipe minus the number caught with the brush. We get $\bar{d} = -0.544$ fish, and $s_D = 0.714$ fish. Because $\bar{d} < 0$, the data suggests that the pipe is less effective at attracting fish than the brush.

a

The statistic for the paired t-test is

$$t = \frac{-0.544 - 0}{0.714/\sqrt{16}} = -3.05$$

The number of degrees of freedom is $16 - 1 = 15$, because we've got 16 pairs of observations. With $\alpha = 0.01$, we'll therefore reject if the test statistic is either above $t_{0.005,15}$ or below $-t_{0.005,15}$. This critical value is 2.947. Because $-3.05 < -2.947$, we can reject the null hypothesis; if the two types of attractors were equally effective, it's not that likely we'd see this big a difference.

b

If we treat this as a two-sample test, then we need s_1, s_2, and $\overline{x}_1$ and $\overline{x}_2$. The first two are given to us by the problem. We could calculate the other two from the data, but notice that $\overline{x_1} - \overline{x_2} = \overline{(x_1 - x_2)}$, which we know from part (a) is -0.544. So, treating this as a two-sample problem,

$$t = \frac{-0.544 - 0}{\sqrt{\frac{(2.48)^2}{16} + \frac{(2.91)^2}{16}}} = -0.57$$

To evaluate this, we need the number of degrees of freedom. But there's a shortcut here (which no one will be penalized for not taking). Looking at the table in appendix A8, you can see that, no matter how many degrees of freedom we have, the area to the right of $|t|$, under the distribution curve, is at least 0.274. So, by symmetry, the area to the left of t at least 0.274, too. The two-sided p-value is sum of these probabilities, which is to say it's at least 0.548. This is not significant by any reasonable standard.

8.68 (Extra Credit)

Looking at problem 55, we see that the question is whether the true mean activation time of sprinklers is equal to 25 seconds (null hypothesis) or greater than 25 seconds (alternative hypothesis). This calls for a one-sided test, which is what Minitab has done. The p-value is 0.043. This is < 0.05, so we can reject the null at the 5% level, but not at the 1% level ($0.043 > 0.01$). So the evidence that the mean activation time is longer than it should be, while decent (we can reject at the 5% level), it isn't conclusive (we can't reject at the stricter 1% level).

8.75 (Extra Credit)

Our null hypothesis is that λ, the average number of weekly requests, $= \lambda_0 = 4.0$. We want to know whether the true λ is greater than this, so we want a one-sided test; the alternative is $\lambda > 4.0$. Our test statistic is

$$z = \frac{\bar{x} - \lambda_0}{\sqrt{\lambda_0 / n}}$$

and we will reject if z is greater than $z_\alpha = z_{0.02} = 2.05$. (Note that we use z_α because it's a one-sided test, not $z_{\alpha/2}$ as in a two-sided test.)

To calculate z, we need $\bar{x}$ and n. We're told that there were 160 requests over 36 weeks, and we're interested in requests per week, so $n = 36$ and $\overline{x} = 160/36 = 4.44$. Thus $z = 1.33 < 2.05$, and we accept the null; there's no strong evidence that the average number of requests per week is over 4.

9.7 (Extra Credit)

Write Δ for the true average amount by which men are more easily bored than women are. The null hypothesis is that $\Delta = 0 = \Delta_0$, and the alternative of interest is that $\Delta > 0$ (that is, men are, on average, more easily bored than women, on average). Because the sample sizes from the two populations are fairly large, we can use a z-test here; it's one-sided, because we've got a one-sided alternative. The test statistic is

$$z = \frac{\overline{x_M} - \overline{x_F} - \Delta_0}{\sqrt{\frac{s_M^2}{n_M} + \frac{s_F^2}{n_F}}} = \frac{10.40 - 9.26 - 0}{\sqrt{\frac{(4.83)^2}{97} + \frac{(4.68)^2}{148}}} = 1.83$$

The p-value (one-sided, remember) is 0.033. Since this is less than $\alpha = 0.05$, we reject the null — there is reasonably solid evidence here that men have a systematically higher average boredom proneness than women do.

9.32 (Extra Credit)

Let's write Δ for the amount by which the true average stress limit of oak exceeds that of fir. The null hypothesis is that $\Delta = \Delta_0 = 1\text{MPa}$. The alternative is that $\Delta > \Delta_0$. We therefore want a one-sided test. Since the population variances are unknown, and the sample sizes are small (14 and 10), but we're willing to assume the individual observations are Gaussian, we'll use a t-test.

$$t = \frac{\overline{x_1} - \overline{x_2} - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{8.48 - 6.65 - 1}{\sqrt{\frac{(0.79)^2}{14} + \frac{(1.28)^2}{10}}} = 1.82$$

Because this is a t-test, we need to know the number of degrees of freedom. The formula is

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}} = 13.85$$

which we round up to 14. With 14 degrees of freedom, the probability of getting a T score of at least 1.82 is 0.046. So we reject the null at any α greater than this. In particular, we would reject the null at the standard 5% level. It's not very likely that we'd observe this big a difference in the breaking strain just by chance.