EEB 596z, Problem Set Two Solutions

1 : Suppose you are studying the number of visitations of a pollinator to a flower. Your hypothesis is that yellow flowers are better than red flowers (in terms of pollinator attraction). Previous studies have found that the number of visitors to red flowers follows a normal distribution with a mean of 200 visits per flower and a variance of 50. Suppose in a sample of 20 yellow flowers that the mean number of visits is 202 with a known variance (of visits per flower) of 50. Again, assume the number of visitors is normally distributed.

(a) What is the probability of this data under the null hypothesis (that yellow and red flowers are equivalent)?

Let $T = 202$ be the mean number of visits. The mean and variance (under the null hypothesis) are $\mu_0 = 200$, and $\sigma_0^2 = 50/20$ (so that $f_0^2 = 50$). Under the null

$$\frac{T-200}{\sqrt{50/20}} \sim U, \qquad \textbf{implying} \quad P\left(U > \frac{202-200}{\sqrt{50/20}}\right) = P\left(U > 1.265\right) = 0.103$$

(b) What is the critical value for a (two-sided) test of the null hypothesis at the $\alpha = 0.05$ level?

$$T_c(0.05) = \mu_0 \pm z_{(1-0.05/2)}\sqrt{50/20} = 200 \pm 1.96\sqrt{50/20} = 196.9, 203.1$$

(c) What are the values for (a) and (b) when the variance for yellow flowers (50) is instead a SAMPLE variance (i.e., an estimate of the true variance)? Hint: Would you now use a normal or a t distribution?

$$\frac{T-200}{\sqrt{50/20}} \sim t_{19}, \qquad \textbf{implying} \quad P\left(t_{19} > \frac{202-200}{\sqrt{50/20}}\right) = P\left(t_{19} > 1.265\right) = 0.1105991$$

in R, use `T<-(202-200)/sqrt(50/20) 0.5; 1-pt(T,19)`.

To find $t_{19,0.975}$ such that $P(t_{19} > t_{19,0.95}) = 0.5$, use (in R) qt(0.975,19) which returns 2.09. Hence, the critical value becomes

$$T_c(0.05, 19) = \mu_0 \pm t_{19,0.95}\sqrt{50/20} = 196.69, 203.31$$

(d) Suppose that yellow flowers are indeed better. Given the sample size (20) and assuming the variance (50) is the true value, how small an effect can we detect using a (two-sided) test of significance of $\alpha = 0.05$ with 80% power?

The power is

$$\Pr\left(T > 203.1\right) + \Pr\left(T < 196.9\right)$$

$$\Pr\left(\frac{t-\mu_1}{\sigma_1} > \frac{203.1-\mu_1}{\sigma_1}\right) + \Pr\left(\frac{t-\mu_1}{\sigma_1} < \frac{196.91-\mu_1}{\sigma_1}\right) =$$

$$\Pr\left(U > \frac{203.1-\mu_1}{\sqrt{50/20}}\right) + \Pr\left(U < \frac{196.91-\mu_1}{\sqrt{50/20}}\right) = 0.80$$

We can solve this itervatively using R. Trying a starting value of 203, in R we can code this as

```
> u1 <-203;
1-pnorm((203.1-u1)/sqrt(50/20)) + pnorm((196.91-u1)/sqrt(50/20))
```

which returns `0.4748`. Trying various values shows that taking $\mu_1 = 204.45$ returns a value of 0.803.

(e) Repeat the calculation in (d) assuming that the variance (50) is now an estimated value, not neccesarily the true value.

We now use the t distribution,

$$\Pr\left((t_{19} > \frac{203.31-\mu_1}{\sqrt{50/20}}\right) + \Pr\left((t_{19} < \frac{196.69-\mu_1}{\sqrt{50/20}}\right) = 0.80$$

In R, we can solve for this using (as $\mu_1 = 204$ as a trail value):

```
> u1 <-204;
1-pt((203.31-u1)/sqrt(50/20),19)+pt((196.69-u1)/sqrt(50/20),19)
```

which returns 0.6664. Using $\mu_1 = 204.65$ returns 0.7964.

(f) Suppose the true mean and variance for yellow flowers are 201 and 10. How large a sample size is required to have a power of 80 percent of detecting a difference between red and yellow using a test of significance with level $\alpha = 0.05$? Compute this for both the normal (variance assumed know) and t (variance estimated) settings.

Variance assumed known (Normal distribution)

Applying Equation A5.4b and recalling for 80 percent power, we use $z_{(1-\beta)} = z_{(0.2)} = 0.842$ and likewise for $\alpha = 0.05$, $z_{(1-\alpha)} = 1.645$,

$$n = \left(\frac{z_{(1-\beta)}\, f_1 + z_{(1-\alpha)}\, f_0}{\mu_1 - \mu_0}\right)^2 = \left(\frac{0.842\sqrt{10} + 1.645\sqrt{50}}{1}\right)^2 = 204$$

Variance estimated (t distribution), here values for $z_{1-\alpha}$ are replaced with $t_{19,1-\alpha}$,

$$n = \left(\frac{t_{(19,1-\beta)}\, f_1 + t_{(19,1-\alpha)}\, f_0}{\mu_1 - \mu_0}\right)^2 = \left(\frac{0.861\sqrt{10} + 1.729\sqrt{50}}{1}\right)^2 = 223.5$$

(g) If the true variance for yellow is 35, what is the probability that we observe a sample variance of 50 (or larger) given our sample size of 20.

Recalling that $\sum^n (x - \overline{x}) \sim \sigma^2\chi^2_{n-1}$, we have

$$\textbf{Var} = \frac{1}{19}\sum (x - \overline{x}) \sim \frac{1}{19}\sigma^2\chi^2_{19} = 1.842\,\chi^2_{19}$$

Hence

$$\Pr(\textbf{Var} > 50) = \Pr\left(1.842\,\chi^2_{19} > 50\right) = \Pr\left(\chi^2_{19} > 27.14\right) = 0.10$$

In R, `1-pchisq(27.14, 19)`