STAT 513 fa 2019 hw 3

1. The following data resulted from an experiment studying the effect of nitrogen fertilizer on lettuce. Data are taken from Kuehl (2000).

Amount of fertilizer	Heads of lettuce per plot			
0 lbs/acre	104	114	90	140
200 lbs/acre	131	148	154	163

Assume that the number of heads of lettuce in the plots is approximately Normally distributed.

Read the data into R with

```
ctrl <- c(104,114,90,140)
trt <- c(131,148,154,163)
```

(a) Let σ_1^2 represent the variance of the number of heads of lettuce under 0 lbs/acre of fertilizer and let σ_2^2 represent that under 200 lbs/acre. For each of the following sets of hypotheses get the p-value based on the data in the table.

i. H_0: $\sigma_1^2 \leq \sigma_2^2$ versus H_1: $\sigma_1^2 > \sigma_2^2$

We have $S_1 = \texttt{sd(ctrl)} = 21.10292$ and $S_1 = \texttt{sd(trt)} = 13.49074$, so that $S_1^2/S_2^2 = 2.446886$. The p-value is the smallest value of α such that we reject H_0, that is, the smallest value of α such that $2.446886 > F_{3,3,\alpha}$. We get

$$\inf\{\alpha : 2.446886 > F_{3,3,\alpha}\} = \texttt{1-pf(2.446886,3,3)} = 0.2408382,$$

which is the area under the pdf of the $F_{3,3}$ distribution to the right of 2.446886.

ii. H_0: $\sigma_1^2 \geq \sigma_2^2$ versus H_1: $\sigma_1^2 < \sigma_2^2$

The p-value is

$$\inf\{\alpha : 2.446886 < F_{3,3,1-\alpha}\} = \texttt{pf(2.446886,3,3)} = 0.7591618,$$

which is the area under the pdf of the $F_{3,3}$ distribution to the left of 2.446886.

iii. H_0: $\sigma_1^2 = \sigma_2^2$ versus H_1: $\sigma_1^2 \neq \sigma_2^2$

The p-value is

$$\begin{aligned}\inf\{\alpha : 2.446886 < F_{3,3,1-\alpha} \text{ or } 2.446886 > F_{3,3,\alpha}\} &= \texttt{2*(1-pf(2.446886,3,3))} \\ &= 0.4816763,\end{aligned}$$

which is the two times the area under the pdf of the $F_{3,3}$ distribution to the right of 2.446886.

(b) Do you believe that the variances σ_1^2 and σ_2^2 are different? Why/why not?

> The evidence is not strong against H_0: $\sigma_1^2 = \sigma_2^2$, so we can probably treat them as though they were equal.

(c) Let μ_1 represent the mean number of heads of lettuce per plot under 0 lbs/acre of fertilizer and let μ_2 represent that under 200 lbs/acre. For each of the following sets of hypotheses get the p-value based on the data in the table, assuming that $\sigma_1^2 = \sigma_2^2$.

i. H_0: $\mu_1 \leq \mu_2$ versus H_1: $\mu_1 > \mu_2$

> We get
>
> $$S^2_{\text{pooled}} = \texttt{(3*var(ctrl) + 3*var(trt))/6} = 313.6667,$$
>
> so
>
> $$\frac{\bar{X}_1 - \bar{X}_2}{S_{\text{pooled}}\sqrt{1/n_1 + 1/n_2}} = \texttt{(mean(ctrl)-mean(trt))/sqrt(313.6667*(1/4+1/4))}$$
> $$= -2.95449.$$
>
> The p-value is the area under the pdf of the t_6 distribution to the right of -2.95449, which is
>
> $$\texttt{1 - pt(-2.95449,6)} = 0.9872679.$$

ii. H_0: $\mu_1 \geq \mu_2$ versus H_1: $\mu_1 < \mu_2$

> The p-value is the area under the pdf of the t_6 distribution to the left of -2.95449, which is
>
> $$\texttt{pt(-2.95449,6)} = 0.0127321.$$

iii. H_0: $\mu_1 = \mu_2$ versus H_1: $\mu_1 \neq \mu_2$

> The p-value is two times the area under the pdf of the t_6 distribution to the right of $|-2.95449|$, which is
>
> $$\texttt{2*(1-pt(2.95449,6))} = 0.02546421.$$

(d) Do you believe the means μ_1 and μ_2 are different? Why/why not?

> There is somewhat persuasive evidence that $\mu_1 < \mu_2$, since the p-value for this test was 0.0127321. This means we would reject H_0: $\mu_1 \geq \mu_2$ at all significance levels greater than 0.0127321.

2. Suppose that in a poll of 743 registered voters, 350 say they will vote for candidate A. Suppose p is the proportion of all registered voters who will vote for candidate A.

   (a) Find the p-values for the following sets of hypotheses using the large-sample tests:

      i. H_0: $p \leq 1/2$ versus H_1: $p > 1/2$

      We get

      `sqrt(743)*(350/743 - 1/2)/sqrt(1/2*(1/2))` $= -1.577517$,

      so the p-value is

      `1-pnorm(-1.577517)` $= 0.9426617$.

      ii. H_0: $p \geq 1/2$ versus H_1: $p < 1/2$

      The p-value is

      `pnorm(-1.577517)` $= 0.05733831$.

      iii. H_0: $p = 1/2$ versus H_1: $p \neq 1/2$

      The p-value is

      `2*(1-pnorm(1.577517))` $= 0.1146766$.

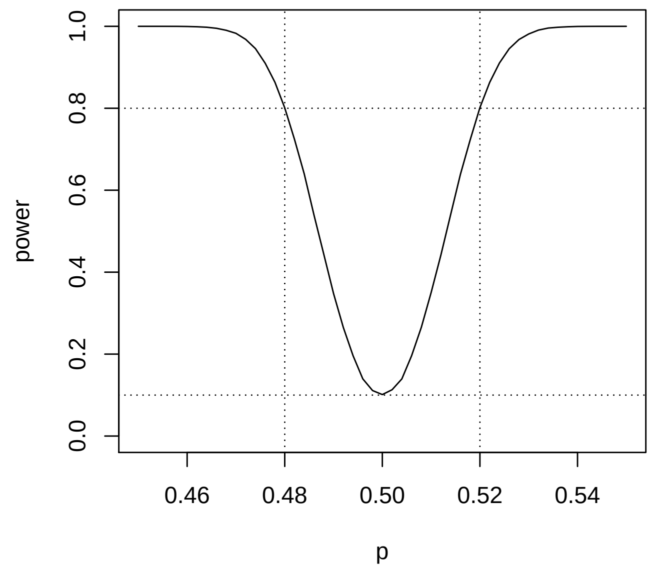
   (b) Suppose a pollster wishes to detect a deviation in p from $1/2$ as small as 0.03 in either direction with probability at least 0.90 while using a test with size $\alpha = 0.05$. What sample size should the pollster take?

   We have

   $$\left[\sqrt{(0.5 \pm 0.03)(1-(0.5 \pm 0.03))}z_{0.10} + \sqrt{0.5(1-0.5)}z_{0.025}\right]^2 / (0.03)^2 = 2914.572,$$

   so take $n^* = 2915$.

   (c) For what sample size does the two-sided test have the power curve depicted below?
   Hint: Identify α, δ^*, $\gamma *$.



The size is $\alpha = 0.10$. The power is greater than $\gamma^* = 0.80$ for all p at a distance of $\delta^* = 0.02$ or more from $1/2$. We have

$$\left[\sqrt{(0.5 \pm 0.02)(1-(0.5 \pm 0.02))}z_{0.20} + \sqrt{0.5(1-0.5)}z_{0.05}\right]^2 / (0.02)^2 = 3862.005,$$

so $n^* = 3863$.

3. Refer to the data in Question 1. Mr. Biogemüsebauer does not wish to fertilize his lettuce plots, but he wants to know if he can expect to get, without fertilizer, a mean yield of more than 100 heads of lettuce per plot (in plots of the same size as those used in the experiment). Otherwise, he will choose to use fertilizer. Let μ be the mean number of heads of lettuce which grow per plot without fertilizer.

   (a) Give the null and alternate hypotheses which are of interest to Mr. Biogemüsebauer.

   H_0: $\mu \leq 100$ versus H_1: $\mu > 100$.

   (b) Give the p-value for testing these hypotheses using the data from Question 1.

   `1-pt(sqrt(4)*(mean(ctrl) - 100)/sd(ctrl),3)` $= 0.1690176$.

   (c) Suppose Mr. Biogemüsebauer wishes to conduct his own experiment using several lettuce plots. If the true mean lettuce yield without fertilizer is 110 heads per plot or more, he wishes to have a Type II error probability no greater than 0.10 when using a test with size 0.05. How many plots of lettuce should he use in his experiment? In your calculations, use the estimated variance for σ^2.

He should use

$$\texttt{ceiling(var(ctrl)*(qnorm(.9) + qnorm(.95))**2/10**2)} = 39$$

plots of lettuce.

References

Kuehl, R. O. (2000). *Designs of experiments: statistical principles of research design and analysis*. Duxbury Press.