Homework Due November 27 (Monday)

p. 284: 6.10, 6.14

p.296: 6.20, 6.22

Chapter 6 Inferences Comparing Two Population Central Values

• The general topic is inference regarding the centers of two possibly different populations

• Objectives:

1. estimate the difference in centers (population means, or medians)

2. test hypotheses about the differences in population means, or medians. For example, test $H_0 : \mu_1 - \mu_2 = 0$ versus $H_a : \mu_1 - \mu_2 > 0$

• We consider two situations:

1) **Two Independent Samples**

Example: Does seeding clouds increase the mean amount of rainfall falling from a seeded cloud? On each of 52 days that were deemed suitable for seeding, and random number table was used to detemine whether or not to seed. A plane was loaded with either NaI, or an inert salt, and the salt was seeded into a clould. Pilots did not know what kind of salt was dropped. The amount of rainfall was measured by radar.

Data: 26 measurements on rainfall when clouds *were not* seeded (sample from population 1), and 26 measurements on rainfall when clouds *were* seeded (sample from population 2).

2) **A Sample of Pairs**

Example: Among right-handed men, is right eye vision better than left eye vision? Measure vision in both eyes for a sample of 30 right-handed men. μ_1 and μ_2 represent the mean vision in left and right eyes of right-handed men, respectivily.

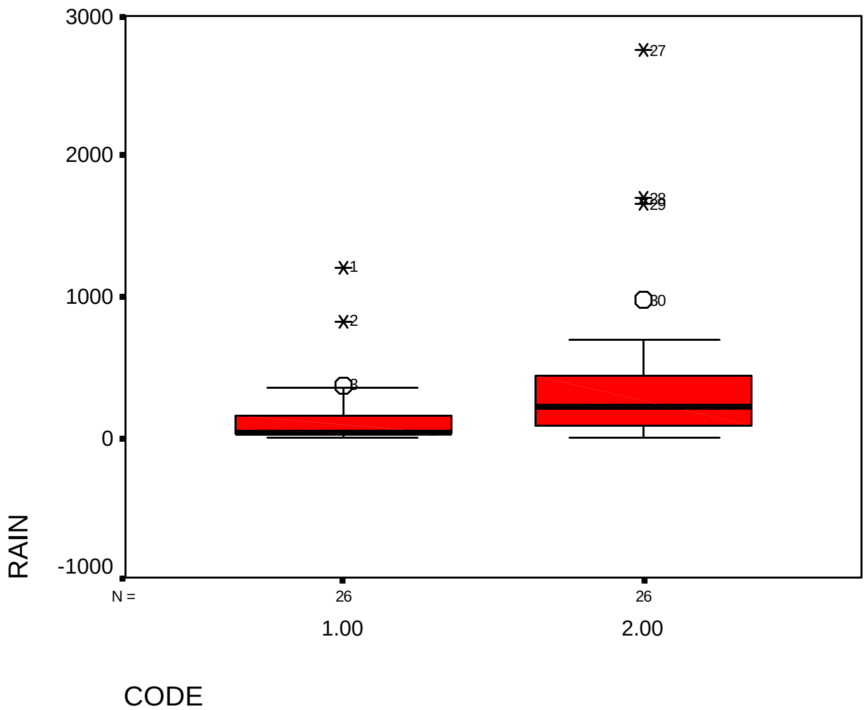
• Samples are *not independent* because men with poor vision in one eye usually have poor vision in the other. Knowing the vision measurement in the right (or left) eye is provides some information about vision in the other. Hence, right- and left-eye vision is not independent.

• Analyzing the differences, right eye score $-$ left eye vision, will usually produce a test statistic with greater power than a test statistic that uses the sample means from each group

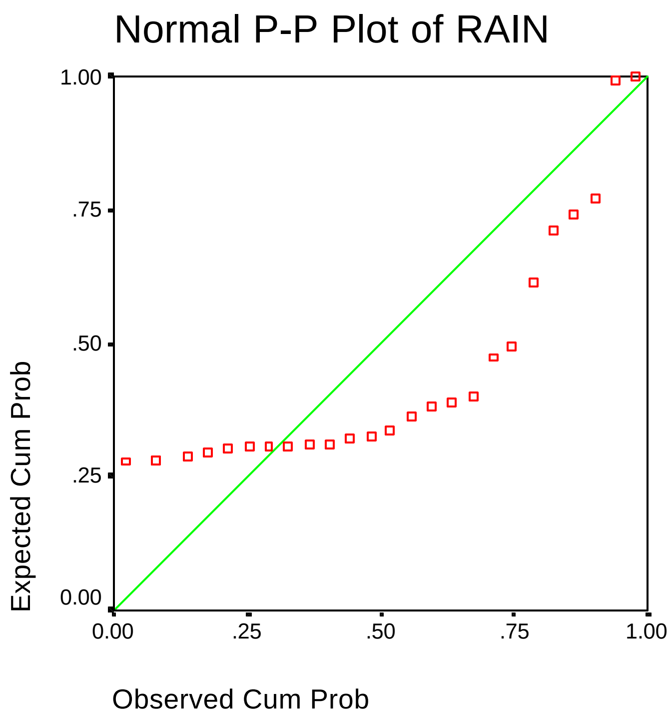
1. Two Independent Samples

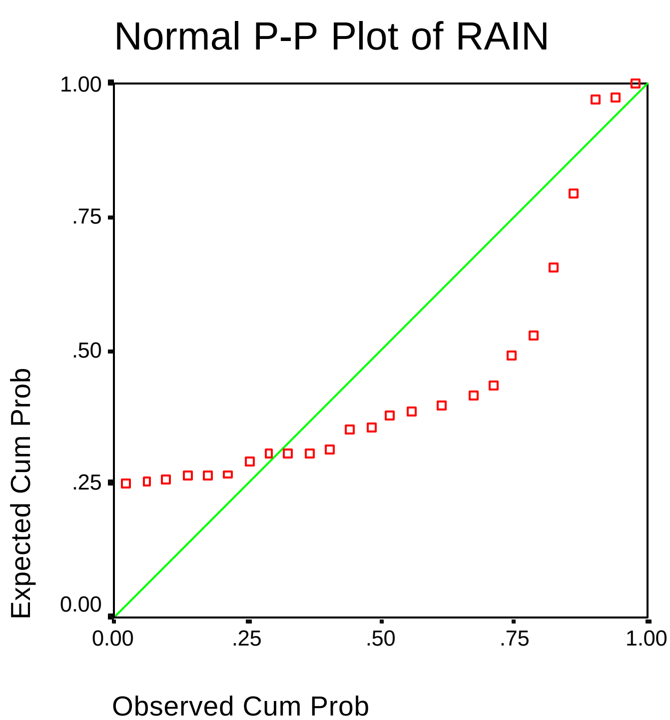
Case Study Simpson, J. Olsen, A., and Eden, J. 1975. "A Bayesian analysis of a multiplicative treatment effect in weather modification," *Technometrics*, **17**, 161-166, and Ramsey, F.L. and Schafer, D.W., 1997, *The Statistical Sleuth*, Duxbury Press, p. 54.

- The cloud seeding data are summarized in the following boxplot.



• Probability-probability plots. These data are from possibly different populations, so we construct P-P plots from both samples. Do not pool the data.

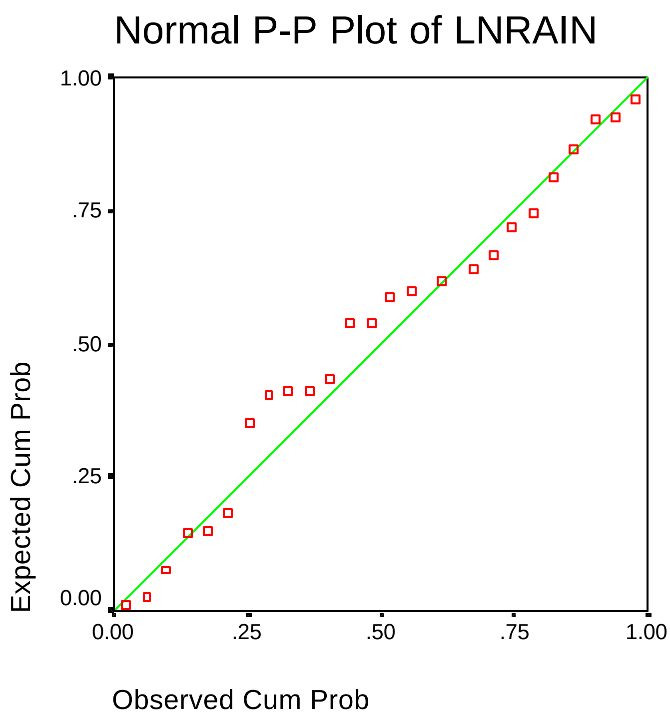




• Skewness sometimes can be removed by a nonlinear transformation. For example, rainfall amounts can be transformed via the natural logarithm. If r_i is the rainfall on the ith day, then,

$$l_i = \ln(r_i + 1) = \log_e(r_i + 1),\ i = 1, \ldots, n$$

are the log-transformed data. The figure below is a P-P plotfor the seeded group



• The log-rainfall medians are $M_1 = 3.79$ (unseeded) and $M_2 = 5.40$ (seeded). Units are log-acre-feet

• Converting back from natural logs to the original scale yields $e^{3.79} = 44.2$ and $e^{5.4} = 221.6$ acre-feet (same as the medians on the original scale).

• Seeding results in an estimated increase of $221.6/44.2 = 5$ times as much rainfall as not seeding. A formal test shows that there is strong evidence that seeding increases the amount of rainfall. Causation is attributable to seeding because this was a controlled experimentation that used randomization

The Basis for Inferential Procedures Regarding $\boldsymbol{\mu_1 - \mu_2}$

• Suppose that $X_1 \sim N(\mu_1, \sigma_1)$ and $X_2 \sim N(\mu_2, \sigma_2)$ are independent random variables.

• σ_1 and σ_2 are the standard deviations of population 1 and 2 respectively

• Two important results are:

1) $$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}),$$

• Note that the standard deviation of $X_1 + X_2$ is $\sqrt{\sigma_1^2 + \sigma_2^2}$

2) $$X_1 - X_2 \sim N(\mu_1 - \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}).$$

• Suppose that $\overline{Y}_1$ and $\overline{Y}_2$ are sample means computed from samples of size n_1 and n_2, respectively. Then,

• $\overline{Y}_1 \sim N(\mu_1, \sigma_1/\sqrt{n_1})$ and $\overline{Y}_2 \sim N(\mu_2, \sigma_2/\sqrt{n_2})$

• $\overline{Y}_1 - \overline{Y}_2 \sim N\Big(\mu_1 - \mu_2, \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}\Big).$

• We use $\overline{Y}_1 - \overline{Y}_2$ as an **estimator of** $\mu_1 - \mu_2$

• To carry out tests regarding $\mu_1 - \mu_2$ and construct CI's for $\mu_1 - \mu_2$, we need to estimate the variablility of the difference of sample means

• The standard deviation of the distribution of $\overline{Y}_1 - \overline{Y}_2$ is

$$\sigma_{\overline{y}_1 - \overline{y}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

• $\sigma_{\overline{y}_1 - \overline{y}_2}$ is also called the **standard error** of the distribution of $\overline{Y}_1 - \overline{Y}_2$

• If we can justify the assumption that the two populations have the same standard deviation, so that

$$\sigma_1 = \sigma_2 = \sigma,$$

then there is a simpler expression for the standard error, namely,

$$\sigma_{\bar{y}_1 - \bar{y}_2} = \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

- $\sigma_{\bar{y}_1 - \bar{y}_2}$ usually is not known, so we estimate $\sigma_{\bar{y}_1 - \bar{y}_2}$ using

$$\widehat{\sigma}_{\bar{y}_1 - \bar{y}_2} = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where

$$\begin{aligned} s_p &= \sqrt{\frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2}} \\ &= \sqrt{\frac{\sum_{i=1}^{n_1} (Y_{i,1} - \overline{Y}_1)^2 + \sum_{i=1}^{n_2} (Y_{i,2} - \overline{Y}_2)^2}{n_1 + n_2 - 2}} \end{aligned}$$

is called the *pooled sample standard deviation*.

- s_p is an estimator of σ, the *common* standard deviation of the two populations

6.2 Inferences about $\boldsymbol{\mu_1 - \mu_2}$: Independent Samples

- In this section, methods are developed assuming that there are two independent samples from two, possibly different, populations.

- The general question of interest is: do the populations have the same mean?

A Confidence Interval for $\mu_1 - \mu_2$

- Recall that $\overline{Y} \sim N(\mu, \sigma_{\overline{y}})$ led to a $100(1 - \alpha)\%$ CI for μ given by

$$\overline{Y} \pm t_{\alpha/2} \frac{s}{\sqrt{n}},$$

where $\overline{Y}$ and s are computed from a random sample of size n, and $t_{\alpha/2}$ is the $1 - \alpha/2$ percentile of the t-distribution with df $= n - 1$.

• Similarly, $\overline{Y}_1 - \overline{Y}_2 \sim N\left(\mu_1 - \mu_2, \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$ leads to a $100(1-\alpha)$ confidence for $\mu_1 - \mu_2$ given by

$$\overline{Y}_1 - \overline{Y}_2 \pm t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},$$

where $t_{\alpha/2}$ is obtained using df $= n_1 + n_2 - 2$.

• The assumptions of this procedure are:

1) The two populations are normal in distribution

2) $\sigma_1 = \sigma_2$

3) The samples are independent

Example

• $\overline{Y}_1 = 3.990$, $\overline{Y}_2 = 5.134 \Rightarrow \overline{Y}_2 - \overline{Y}_1 = 1.144.$

• $\alpha = 0.05,$ df $= 52 - 2 = 50 \Rightarrow t_{\alpha/2} = 2.021.$

• Transform rainfall to natural logarithm of rainfall to reduce the skewness of the sample data. For the transformed data,

$\overline{Y}_1 = 3.990,\ s_1 = 1.642, n_1 = 26,$

$\overline{Y}_2 = 5.134,\ s_2 = 1.600\ \ n_2 = 26.$

• Then, $\overline{Y}_2 - \overline{Y}_1 = 1.144,$

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{25 \times 1.642^2 + 25 \times 1.600^2}{50}}$$
$$= \sqrt{2.628} = 1.621,$$

and

$$\begin{aligned}\widehat{\sigma}_{\,\overline{y}_2-\overline{y}_1} &= s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\\ &= 1.621\sqrt{\frac{1}{26}+\frac{1}{26}}\\ &= 0.4496.\end{aligned}$$

• A 95% CI for $\mu_2 - \mu_1$ is

$$\begin{aligned}\overline{Y}_1 - \overline{Y}_2 \pm t_{\alpha/2}\widehat{\sigma}_{\,\overline{y}_1-\overline{y}_2}&\\ &= 1.144 \pm 2.01 \times 0.4496\\ &= 0.240 \text{ to } 2.048.\end{aligned}$$

• The interval can be written as $(0.240{,}2.048)$ or $[0.240{,}2.048]$ log acre-feet

• This 95% CI is for the natural log of rainfall. To get an approximate 95% interval for the mean difference in rainfall (mm), we reverse the log transformation by computing the anti-logs of $L = 0.240$ and $U = 2.048$

• Computing anti-logs gives an approximate 95% CI for $\mu_1 - \mu_2$:

$$(e^{0.240}, e^{2.048}) = (1.27{,}7.74) \text{ acre-feet}$$

Hypothesis Testing - Two Independent Samples

• In this section, a testing procedure is develolped for the difference in the population means μ_1 and μ_2

• Specifically, we hypothesize that the difference is a number denoted by D_0 (the null hypothesis difference). Commonly, D_0 is 0

• We test $H_0 : \mu_1 - \mu_2 = D_0$ versus one of three alternatives:

1) $H_a : \mu_1 - \mu_2 > D_0$,

2) $H_a : \mu_1 - \mu_2 < D_0$, or

3) $H_a : \mu_1 - \mu_2 \neq D_0$

• The assumptions of this procedure are:

1) The two populations are normal in distribution

2) $\sigma_1 = \sigma_2$

3) The samples are independent

• The test statistic is

$$T = \frac{\overline{Y}_1 - \overline{Y}_2 - D_0}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}.$$

• If H_0 is true, then $T \sim \mathcal{T}_{n_1+n_2-2}$.

• This is called the *pooled two-sample* T test. There is a similar test called the *separate-variance* T test.

• The pooled two-sample T statistic is used if $\sigma_1 = \sigma_2$ is reasonable; the separate-variance statistic T is used if the assumption cannot be justified

• For now, the discussion is limited to the *pooled two-sample* T statistic

• The rejection region depends on the alternative hypothesis. For df $= n_1 + n_2 - 2$, look up t_α or $t_{\alpha/2}$ for a one- or two-sided test, respectively,

• The rejection regions for each of the three alternative above are

1) All values of T that are greater than t_α

2) All values of T that are less than $-t_\alpha$

3) All values of T that are less than $-t_{\alpha/2}$, and all values of T that are greater than $t_{\alpha/2}$

Example

$H_0 : \mu_2 - \mu_1 = 0$ (i.e., $D_0 = 0)$ versus $H_a : \mu_2 - \mu_1 > 0.$

• Convert mm rainfall to natural logarithm of rainfall. Then,

$$\overline{Y}_1 = 3.990, \overline{Y}_2 = 5.134 \quad \Rightarrow \quad \overline{Y}_2 - \overline{Y}_1 = 1.144$$

and

$$s_1 = 1.642, s_2 = 1.600,\ n_1 = 26,\ n_2 = 26$$

$$\Rightarrow s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{25 \times 1.642^2 + 25 \times 1.600^2}{50}}$$
$$= \sqrt{2.628} = 1.621,$$

- Then,

$$T = \frac{\overline{Y}_2 - \overline{Y}_1 - D_0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{1.144}{1.621\sqrt{\frac{2}{26}}} = 2.544,$$

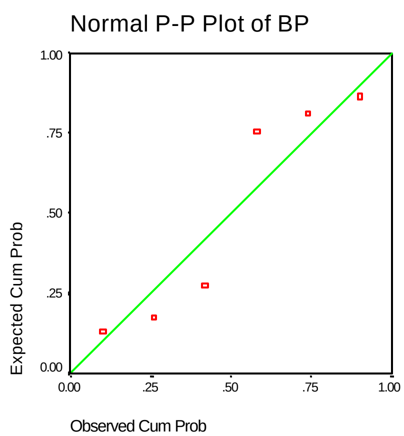
Also, df $= 26 + 26 - 2 = 50$. An exact calculation (use SPSS) gives p-value $= P(T > 2.544) = 0.007$.

- Based on these data, there is very strong evidence that seeding increases the mean amount of rain on days that are considered to be suitable for seeding.

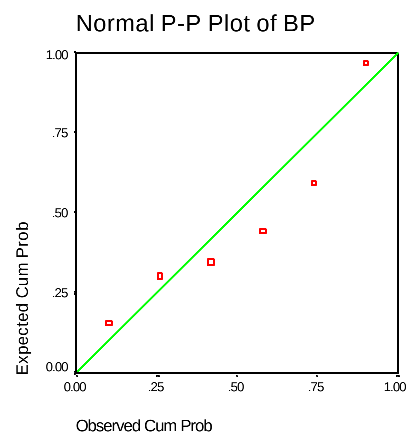
Problem 6.4, p. 280

- Probability-probability plots are used assess the normal distribution assumption. The sample observations indicate that the assumption of normality is difficult to support

26° C



5° C



- The sample means and standard deviations are

$$\overline{Y}_1 = 165.8 \quad s_1 = 14.8$$
$$\overline{Y}_2 = 378.5 \quad s_2 = 23.9,$$

where group 1 is the 26° C group, and group 2 is the 5° C group

• The sample standard deviations quite different, but then again, the sample sizes are small and σ_1 and σ_2 are not well-estimated.

• Levene's test for equality of variances will help clarify whether I can safely assume that $\sigma_1 = \sigma_2$. The hypotheses are $H_0: \sigma_1 = \sigma_2$ and $H_a: \sigma_1 \neq \sigma_2$. The test statistic and p-value measure the strength of evidence against H_0 and in favor of H_a

• Levene's test should be viewed as an informal test because the accuracy of its p-value calculation depends strongly on the assumption that the data are drawn from normal populations

• According to the description of the study, it is reasonable to assume that the samples are independent

• For now, I proceed with caution and use the pooled two-sample t-statistic. Therefore, my hypotheses will be about the population *means* (versus medians)

a) The hypotheses are:

$$H_0: \mu_2 - \mu_1 = 0 \text{ and } H_a: \mu_2 - \mu_1 > 0$$

where μ_1 and μ_2 are the population means for the 26° C group and the 5° C group, respectively

• The test statistic is

$$T = \frac{\overline{Y}_2 - \overline{Y}_1}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}.$$

• The value of s_p is

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{5 \times 14.8^2 + 5 \times 23.9^2}{10}} = 19.88,$$

then,

$$T = \frac{378.5 - 165.8}{19.88\sqrt{\frac{1}{6} + \frac{1}{6}}} = 18.5.$$

• To test at the $\alpha = 0.05$ level, we reject for large values of T. Specifically, if T is larger than the critical value $t_\alpha = 1.812$ (df $= 10$).

• My conclusion is that there is abundant evidence that $\mu_2 - \mu_1 > 0$. The issue of whether to use the pooled- or separate-variance t is irrelevant because the evidence is overwhelming, and re-doing the test will not change the conclusion

b) Assumptions were discussed above.

c) The confidence interval is given by

$$\overline{Y}_2 - \overline{Y}_1 \pm t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where $t_{\alpha/2} = 2.228$ (df $= 10$ and $\alpha/2 = 0.025$). Hence, the CI is

$$378.5 - 165.8 \pm 2.228 \times 19.88\sqrt{\frac{1}{6} + \frac{1}{6}} = (187.1, 238.3)$$

Unequal Population Variances

• Now suppose that it is not true, or that we unwilling to assume that $\sigma_1 = \sigma_2$

• If $n_1 = n_2$, then, t-procedures are O.K. provided that σ_1^2 and σ_2^2 do not differ by more than a factor of 3.

• If $n_1 > 2n_2$ or $n_2 > 2n_1$, then, t-procedures are very sensitive to the assumption of equal variances

• If σ_1^2 and σ_2^2 are substantially different, then the *separate variance* t-statistic is used to test $H_0 : \mu_1 - \mu_2 = D_0$ versus one of three alternatives:

1) $H_a : \mu_1 - \mu_2 > D_0$,

2) $H_a : \mu_1 - \mu_2 < D_0$, or

3) $H_a : \mu_1 - \mu_2 \neq D_0$

- The *separate variance* t-statistic is

$$T' = \frac{\overline{Y}_1 - \overline{Y}_2 - D_0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}.$$

- Note that the denominator differs from that of the pooled variance t-statistic
- Under H_0, $T' \dot{\sim} \mathcal{T}_{\text{df}}$ where

$$\text{df} = \frac{(n_1 - 1)(n_2 - 1)}{(n_2 - 1)c^2 + (n_1 - 1)(1 - c)^2},$$

and

$$c = \frac{s_1^2/n_1}{s_1^2/n_1 + s_2^2/n_2}.$$

- If df is not an integer, then round *down* to nearest integer.
- Use T' in the same manner as T for determining rejection regions and CIs.

Example

- 7 lambs receive a drug against tapeworms, 6 receive no drug. Responses were worm counts in slaughtered lambs after 6 months.
- Let μ_1 denote the mean number of tapeworms for lambs receiving the drug, and let μ_2 denote the mean number of tapeworms for lambs receiving no drug. The sample standard deviations from the two samples are denoted by s_1 and s_2, respectively
- Test: $H_0 : \mu_1 - \mu_2 = 0$ versus $H_a : \mu_1 - \mu_2 < 0$.

$\overline{Y}_1 = 9.00$, $s_1^2 = 38.67$,$n_1 = 7$,

$\overline{Y}_2 = 40.17$, $s_2^2 = 258.2$, $n_2 = 6$.

- Note that $s_2^2 > 3s_1^2$.
- Then, $\overline{Y}_1 - \overline{Y}_2 = 9.00 - 47.17 = -31.17$.

- $$T' = \frac{-31.17}{\sqrt{\dfrac{38.67}{7} + \dfrac{258.2}{6}}} = -4.47.$$

- Also,

$$c = \frac{s_1^2/n_1}{s_1^2/n_1 + s_2^2/n_2} = \frac{38.67/7}{38.67/7 + 258.2/6} = 0.114,$$

$\Rightarrow c^2 = 0.013$, and

$$\text{df} = \frac{(n_1 - 1)(n_2 - 1)}{(n_2 - 1)c^2 + (n_1 - 1)(1 - c)^2} = \frac{6 \times 5}{5 \times 0.013 + 6 \times (1 - 0.114)^2} = 6.28.$$

Use $\text{df} = 6$ and look up a critical value in Table 2.

- If $\alpha = 0.05$, then the rejection region is all values of T' that are small than -1.943

- There is sufficient evidence to reject H_0 and conclude that there is a reduction in mean number of worms between the drug-treated and control populations.

6.3 A Nonparametric Alternative: The Wilcoxon Rank Sum Test

- This statistic is called a distribution-free statistic because no distributional assumptions about the sampled populations are made

- It is equivalent to the Mann-Whitney U-statistic

- We test H_0 : the two populations are exactly the same (identical) versus one of three alternatives:

1) H_a : population 1 is shifted to the right of population 2, but otherwise the same

2) H_a : population 1 is shifted to the left of population 2, but otherwise the same

3) H_a : the populations have different centers. That is, one population is shifted with respect to the other

- **Assumptions**: samples are independent

- We do **not require** assumptions of normality or symmetry
- The t**est statistic** is

$$T = \text{sum of the ranks for the observations from pop'n 1}$$

- Under H_0, the ranks from population 1 samples are randomly distributed among the population 2 ranks

- If $n_1 = 2$ and $n_2 = 4$, then we can find the distribution of T by listing every combination of 2 ranks, and their sum, when drawing 2 integers from the set $\{1,2,3,4,5,6\}$. The probability distribution of T is

t	3	4	5	6	7	8	9	10	11
$P(T=t)$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{2}{15}$	$\frac{3}{15}$	$\frac{3}{15}$	$\frac{2}{15}$	$\frac{1}{15}$	$\frac{1}{15}$

Steps to Computing the Test Statistic

1) Combine the samples and rank the observations from smallest to largest
2) Sum the ranks for the observations from sample 1

Notes

- If H_0 is true , the distribution of T has mean and variance

$$\mu_T = \frac{n_1(n_1+n_2+1)}{2} \text{ and } \sigma_T^2 = \frac{n_1 n_2(n_1+n_2+1)}{12}.$$

- If population 1 is shifted to the right of population 2, then T should be large (larger than μ_T)
- If population 1 is shifted to the left of population 2, then T should be small (smaller than μ_T)
- If population 1 is shifted to the right, or left, of population 2, then T should be substantially larger, or smaller, than μ_T

The Rejection Region

- If $n_1 \leq 10$ and $n_2 \leq 10$, then use Table 5 to find a critical value (T_U or T_L, depending on H_a, for a chosen level of α)

The rejection regions, corresponding to alternatives 1) - 3), are

1) All values of T larger than T_U

2) All values of T less than T_L

3) All values of T larger than T_U and all values of T less than T_L

• If $n_1 > 10$ or $n_2 > 10$, then use a normal approximation. It says that if H_0 is true, then

$$\frac{T - \mu_T}{\sigma_T} \dot{\sim} N(0,1).$$

• So, we can compute approximate p-values in the same manner as with the Z- and t-statistics. That is, the p-value is a tail area to the right (or left) of the test statistic

• For example, if the alternative hypothesis is H_a : population 1 is shifted to the right of population 2, then large values of T are evidence against H_0 and in favor of H_a, so

$$p\text{-value} = P(T \geq t)$$

where t is the observed value of T

• Ties. Sometimes observations have the same value and there is no obvious way to rank them.

• Solution: average the ranks of the ties, and replace the ranks of the ties with the average rank. Example: $S_1 = \{1,3,4,5,9,12\}$, $S_2 = \{2,3,4,6,9\}$

Observation	rank	rank	Population
1	1	1	1
2	2	2	2
3	3^*	3.5	1
3	4^*	3.5	2
4	5^{**}	5.5	1
4	6^{**}	5.5	2
5	7	7	1
6	8	8	2
9	9^{***}	9.5	1
9	10^{***}	9.5	2
12	11	11	1

*tied, **tied, ***tied

• $T = 1 + 3.5 + 5.5 + 7 + 9.5 + 11 = 38.5$

- $$\mu_T = \frac{n_1(n_1+n_2+1)}{2} = \frac{6(6+5+1)}{2} = 36,$$
- $$\sigma_T^2 \doteq \frac{n_1 n_2(n_1+n_2+1)}{12} = \frac{6\times 5(6+5+1)}{12} = 30.$$
- Ott and Longnecker (p. 292) provide an alternate formula for σ_T^2 when there are large numbers of ties
- To finish the example, suppose that the hypotheses are

H_0 : the two populations are exactly the same (identical) versus

H_a : population 1 is shifted to the right of population 2, but otherwise the same

and $\alpha = 0.05$. From Table 5, $n_1 = 6$, $n_2 = 5$ and $T_U = 41$ (one-sided).

$$T = 38.5 < 41 = T_U,$$

so we cannot reject H_0.

- To illustrate the calculation,

$$p\text{-value} = P\left(\frac{T-\mu_T}{\sigma_T} > \frac{38.5-36}{\sqrt{30}}\right) \approx P(Z > .456) = 0.32.$$

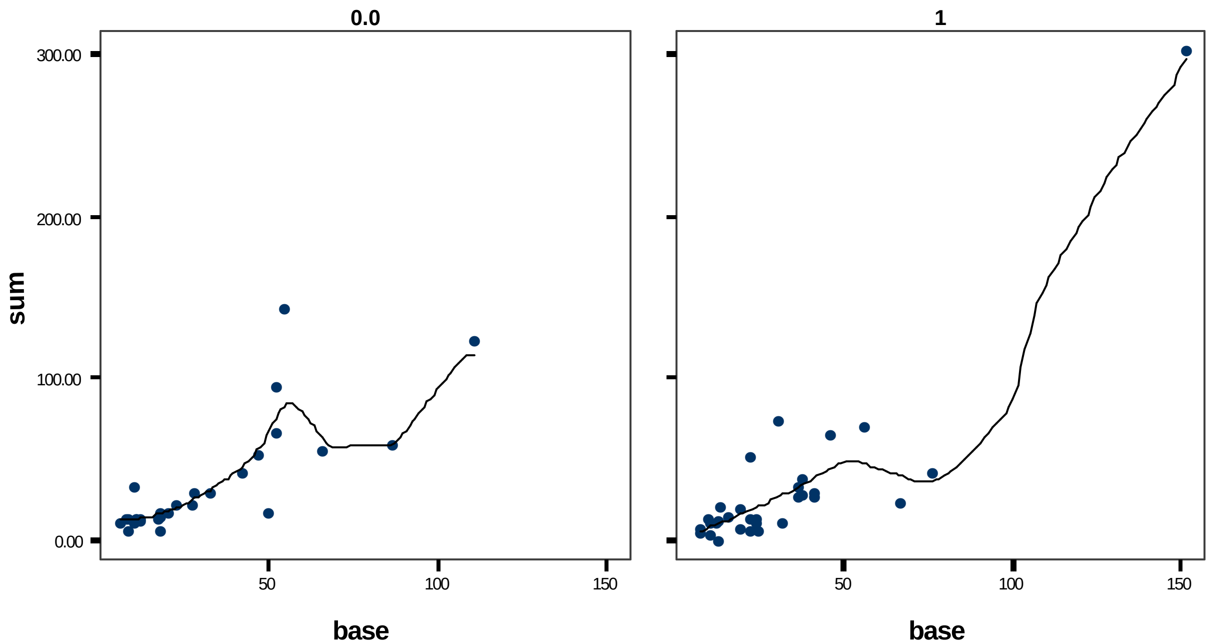
- The normal approximation p-value calculation should be used only if $n_1 > 10$ and $n_2 > 10$, so I have little faith in the computation above

Case Study Thall, P.F. and Vail, S.C. 1990. Some covariance models for longitudinal count data with overdispersion. *Biometrics*, **46**, 657-671. These are *longitudinal* data from a study of the effectiveness of a drug for suppressing seizures in epileptics. 59 subjects were randomly assigned to treatment and placebo groups, and each subject reported the numbers of seizures experiences in each of 4 consecutive two-week periods. Also recorded was age and baseline seizure count.

- Baseline seizure count is important because there are large differences among individuals with respect to seizure rate
- The observations are not independent because there are 4 measurements on each individual. So, I define a new variable which is the sum of the 4 counts and use it instead

Figure. Number of seizures summed over the four observation periods plotted against baseline

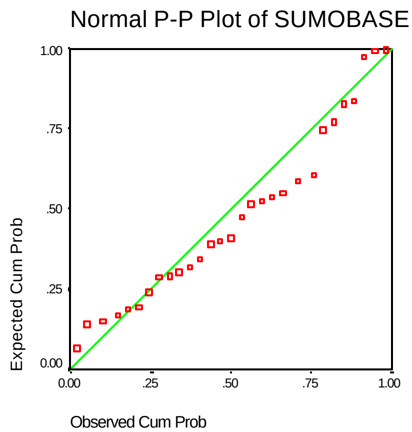
Placebo Group $(n_P = 28)$ Drug Group $(n_D = 31)$



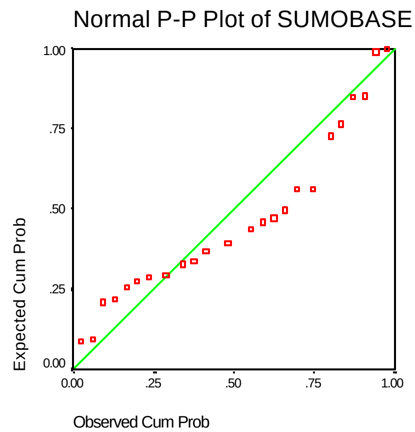
• This plot indicates that there is a lot of variablility associated with the baseline count. One way to partially remove its effect is to scale the counts using the baseline variable. That is, define a new variable given by sum/baseline.

• Normal probability plots for sum/baseline are

Placebo Group



Drug Group



• It is best not to use the t procedures because the normal assumption is not justified. On the other hand, the samples moderate in size, $n_P = 28$ and $n_D = 31$. Moreover, $s_P = 0.631$ and $s_D = 0.575$. So, a t procedure will be approximately correct. Rather than worry anymore, I will use the Wilcoxon Rank Sum test

• The Wilcoxon Rank Sum can be used to test

H_0 : the two populations are exactly the same, versus

H_a : The drug population is shifted to the left (fewer seizures) of the placebo population, but otherwise the same

• If H_a is true, then T should be smaller than μ_T, which is computed assuming H_0 to be true

• I will define the Wilcoxon Rank Sum statistic T to be the sum of the ranks for the drug group. Then, $T = 779$,

• $\mu_T = \dfrac{n_D(n_P + n_D + 1)}{2} = \dfrac{31(28+31+1)}{2} = 930$

$\sigma_T \doteq \sqrt{\dfrac{n_P n_D(n_P + n_D + 1)}{12}} = 65.88$

and

$$P(T \leq 779) \approx P(Z < \frac{779-930}{65.88}) = 0.011.$$

• My conclusion is that there is strong evidence that the drug suppresses seizures (p-value $= 0.011$)

Paired-Data Procedures

• We have already used these procedures for hypothesis testing

• **Example** - (from Ramsey and Schafer 1997, p. 29). Suddath, et al. 1990, "Anatomical abnormalities associated in the brains of monozygotic twins discordant for schizophrenia", New England Journal of Medicine, **322**(12): 789-93, found 15 twins where one was schizophrenic and the other was not. An MRI was used to measure the left hippocampus of the subjects.

• Generally, there are substantial differences among individuals with respect to brain morphology

• Differences between twins respect to brain morphology are much smaller, in general, than the differences among members of the two populations

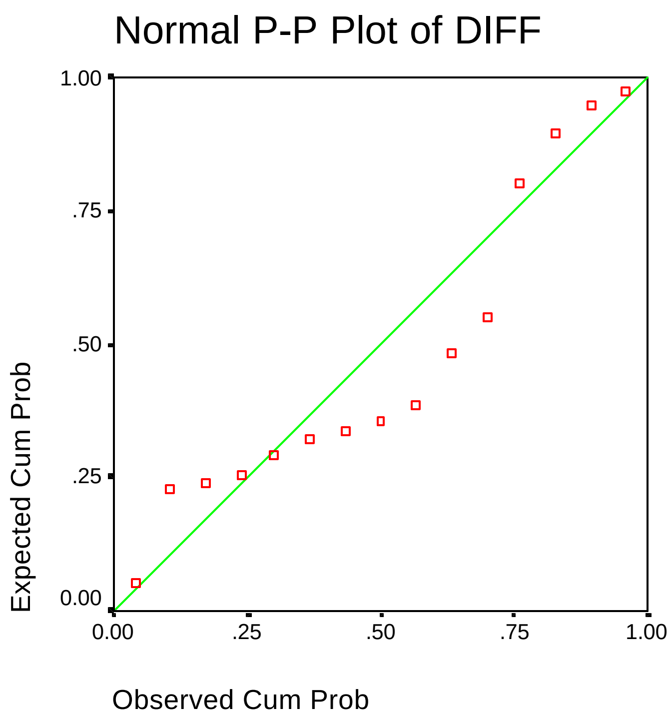
• Paired-data procedures exploit the structure of the data by using the differences for inference, rather than the original data

• Specifically, we convert the data to differences, and carry out the ordinary one-sample t-procedures on the differences

• For example, the differences in volumes (cm^3) of left hippocampus were

Pair	Unaffected	Affected	Difference
1	1.94	1.27	0.67
2	1.44	1.63	−0.19
3	1.56	1.47	0.09
4	1.58	1.47	0.19
5	2.06	1.93	0.13
6	1.66	1.26	0.40
7	1.75	1.71	0.04
8	1.77	1.67	0.10
9	1.78	1.28	0.50
10	1.92	1.85	0.07
11	1.25	1.02	0.23
12	1.93	1.34	0.59
13	2.04	2.02	0.02
14	1.62	1.59	0.03
15	2.08	1.97	0.11

• A normal probability plot of the *differences* does not look sufficiently normal to assume that the t-procedures will achieve the specified α levels



• For now, I will proceed with the t-procedures

• The mean and standard deviation of the differences(unaffected – affected) are

$$\overline{D} = 0.199\,\text{cm}^3,\; s_D = 0.238\,\text{cm}^3.$$

• Let μ_D denote the mean difference (unaffected – affected)between twins in left hippocampus volume. Then, the hypotheses of interest are

• $$H_0 : \mu_D = 0 \text{ versus } H_a : \mu_D \neq 0.$$

and the test statistic is

• $$T = \sqrt{n}\frac{\overline{D} - \mu_D}{s_D} = \sqrt{15}\frac{0.199 - 0}{0.238} = 3.24$$

• To assess the strength of evidence, I will compute the p-value. It is

$$P(|T| > 3.24) = 2P(T > 3.24) = 0.0045.$$

• Conclusion: there is strong evidence that these populations differ with respect to mean left hippocampus volume. Because the evidence is so strong, the lack of conformity of the sample differences to a normal distribution probably does not affect the conclusion, though the p-value is not very accurate

• We cannot say that the difference is caused by schizophrenia

• An approximate $100(1-\alpha)$% CI for the difference μ_D is

$$\overline{D} \pm t_{\alpha/2}\frac{s_D}{\sqrt{n}}.$$

• For these data, an approximate 95% confidence interval

$$\overline{D} \pm t_{.025}\frac{s_D}{\sqrt{n}} = 0.199 \pm 2.145\frac{0.238}{\sqrt{15}} = (0.067, 0.331).$$

• It is not correct so say that this is a 95% confidence interval; it is an *approximate* 95% confidence interval

• Ott and Longnecker summarize these procedures on p. 303

The Wilcoxon Signed Rank Test

• A nonparametric alternative to the paired t-test

• Use when the distribution of differences is symmetric, but not necessarily normal. We can also use the sign test if the original data are replaced with the differences, but the Wilcoxon signed rank statistic generally has greater power

• Test H_0 : the distribution of differences is symmetric about D_0, versus

1). H_a : the distribution of differences is shifted to the right of D_0

2). H_a : the distribution of differences is shifted to the left of D_0

3). H_a : the distribution of differences is shifted to the right, or to the left, of D_0

Computation of the Wilcoxon Signed Rank Test

1. compute pairwise differences and subtract D_0 from every difference (under H_0 this distribution is symmetric about 0)

2. Take the absolute value of the differences, and delete any difference equal to 0

Set $n =$ # of nonzero differences

3. Rank the n absolute differences

4. The test statisitic depends on H_a:

1). H_a : the distribution of differences is shifted to the right of D_0. Then the test statistic is the sum of negative-signed ranks. Call it T_-. *Small* values of T_- support H_a

2). H_a : the distribution of differences is shifted to the left of D_0. Then the test statistic is the sum of positive-signed ranks. Call it T_+. *Small* values of T_+ support H_a

3). H_a : the distribution of differences is shifted to the right, or to the left, of D_0. Then the test statistic is $T = \min\{T_+, T_-\}$. *Small* values of T support H_a

- For example, for the twins data

Difference	Abs. Difference	Rank	Difference Sign
0.67	0.67	15	$+$
-0.19	0.19	9.5	$-$
0.09	0.09	5	$+$
0.19	0.19	9.5	$+$
0.13	0.13	8	$+$
0.40	0.40	12	$+$
0.04	0.04	3	$+$
0.10	0.10	6	$+$
0.50	0.50	13	$+$
0.07	0.07	4	$+$
0.23	0.23	11	$+$
0.59	0.59	14	$+$
0.02	0.02	1	$+$
0.03	0.03	2	$+$
0.11	0.11	7	$+$

$$T = \text{min}\{T_+, T_-\} = \text{min}\{9.5, 108.5\} = 9.5$$

- If $n \leq 50$, compare T to the critical value in Table 6 (p. 1098) for some α, or compute an approximate p-value.

- $T = 9.5$ and $n = 15$ and Table 6 $\Rightarrow$ $.002 < p$-value < 0.005. From SPSS, p-value ≈ 0.004.

- If $n > 50$, compute $\mu_T = \dfrac{n(n+1)}{4}$ and $\sigma_T = \sqrt{\dfrac{n(n+1)(2n+1)}{24}}$ (make adjustment for ties - see Ott and Longnecker, p. 308).

- For H_a : the distribution of differences is shifted to the right of D_0, the test statistic is T_-, and

$$p\text{-value} \approx P(Z \leq \frac{T_- - \mu_T}{\sigma_T})$$

where $Z \sim N(0,1)$.

• For H_a : the distribution of differences is shifted to the left of D_0, the test statistic is T_+, and

$$p\text{-value} \approx P(Z \leq \frac{T_+ - \mu_T}{\sigma_T})$$

where $Z \sim N(0,1)$.

• For H_a : the distribution of differences is shifted to the right, or to the left, of D_0, the test statistic is $T = \min\{T_+, T_-\}$ and

$$p\text{-value} \approx 2P(Z \leq \frac{T - \mu_T}{\sigma_T})$$

where $Z \sim N(0,1)$.

Another Application of the Signed Rank Test for One Population

• Suppose that we wish to test H_0 : the population is symmetrically distributed about D_0, versus

1). H_a : the population is shifted to the right of D_0

2). H_a : the population is shifted to the left of D_0

3). H_a : the population is shifted to the right, or to the left, of D_0

• We can use the Wilcoxon signed rank statistic if we subtract D_0 from every observation and treat these values as if they were differences in the computation of the test statistic

A Key for Identifying the Appropriate Inferential Procedure

A) One population of interest

1) Population is approximately normal in distribution

a) Population standard deviation σ is known (a rare situation). Inferential procedures are based on the Z-statistic:

$$Z = \sqrt{n}\frac{\overline{Y} - \mu_0}{\sigma_{\overline{y}}}$$

Under $H_0 : \mu = \mu_0$, $Z \sim N(0,1)$

b) Population standard deviation σ is *not* known. Inferential procedures are based on the t-statistic. :

$$T = \sqrt{n}\frac{\overline{Y} - \mu_0}{s_{\overline{y}}}$$

Under H_0, $T \sim \mathcal{T}_{\text{df}}$, where $\text{df} = n - 1$. If n is larger than 50, then, the Z table can be used to find p-values and rejection regions.

2) Population is not normal in distribution

a) Transform the original data (e.g., the natural logarithm) to get normality

b) No transformation can be found. Use the Wilcoxon Signed Rank test if the population is symmetrically distributed

c) No transformation can be found, and the population does not appear to be symmetrically distributed. Use the Sign test

B) Two populations of interest

1) Independent samples.

a) Populations are both approximately normal in distribution

i) $\sigma_1^2 = \sigma_2^2$. Use pooled variance t-procedures

ii) $\sigma_1^2 \neq \sigma_2^2$. Use separate-variance t-procedures

- The separate-variance t-procedure is a conservative test. You should use it if it is not clear that $\sigma_1^2 = \sigma_2^2$

b) One or both populations are not normal in distribution

i) Try to find transformations to normality

ii) No transformation can be found. Use the Wilcoxon Rank Sum test

2) Paired data (observations are not independent)

a) The population of *differences* is approximately normal. Use the paired data t-procedures

a) The population of *differences is not* normal in distribution. Use the Wilcoxon Signed Rank Test

- References on nonparametrics

1. Lehman, E.L. 1975. *Nonparmetrics*. Wiley and Sons.

2. Sprent, P. 1989. *Applied Nonparametric Statistical Methods.* Chapman and Hall.

3. Conover, W.J. 1980. *Practical Nonparametric Statistics*, 2nd Ed. Wiley and Sons.