Chapter 11

Inferences on Two Samples

11.1 Inference about Two Population Proportions

We look at the test and confidence interval for difference in two proportions, *in*dependent samples, $p_1 - p_2$,

$$z_0 = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}}, \qquad \hat{p}_1 - \hat{p}_2 \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}},$$

where we assume large simple random samples are chosen and $n_1\hat{p}_1(1 - \hat{p}_1) > 10$, $n_2\hat{p}_2(1 - \hat{p}_2) > 10$. If sampled from finite populations, $n_1 \leq 0.05N_1$, $n_2 \leq 0.05N_2$. For test, sample proportions are pooled

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}.$$

We also look at differences in proportions for matched data (*always* a two-sided test) and at sample size calculations.

Exercise 11.1 (Inference about Two Population Proportions)

1. *Independent or dependent?*
   Identify whether the following samples are independent or dependent.

   (a) **independent / dependent**
   407 male doctors chosen at random from the military are compared to 7363 male doctors chosen at random from civilian life

   (b) **independent / dependent**
   for a smoking study, 958 males chosen at random are compared to 869 females chosen at random

(c) **independent / dependent**
blood is drawn from a female and split into a control vial and a treatment vial; the same is done for 3 other females; the control blood is compared to the treatment blood

(d) **independent / dependent**
10 cows are given *both* control feed and "gentech" feed (at different times), their responses to these two treatments are measured and then compared

(e) **independent / dependent**
blood is drawn from 4 females chosen at random and compared to blood drawn from a different 4 females also chosen at random

(f) **independent / dependent**
plasma is drawn from 9 males chosen at random and compared to plasma drawn from 6 females also chosen at random

(g) **independent / dependent**
returns from 10 small companies chosen at random are compared to returns from 10 large companies also chosen at random

(h) **independent / dependent**
354 patients are given *both* drug and placebo (at different times), their responses to these two treatments are measured and then compared

Sampling is *dependent* if individuals in one sample are used to determine individuals in other sample.

2. *Inference $p_1 - p_2$, large independent samples: doctors.*
Consider number of male doctors in military and civilian hospitals. Test claim there is a *smaller* proportion of male doctors in military than in civilian life at $\alpha = 0.05$. Calculate 95% CI of difference in proportions.

	military (1)	civilian (2)
male doctors	358	6786
total doctors	407	7363

(a) *Hypothesis test.*

i. *Check assumptions.*
Since $n_1\hat{p}_1(1-\hat{p}_1) = 407 \cdot \frac{358}{407}\left(1 - \frac{358}{407}\right) \approx 43.1 > 10$, and
$n_2\hat{p}_2(1-\hat{p}_2) = 7363 \cdot \frac{6786}{7363}\left(1 - \frac{6786}{7363}\right) \approx 531.8 > 10$,
assumptions **are** / **are not** satisfied, so continue.

ii. *Statement.* Choose one.

A. $H_0 : p_1 - p_2 = 0$ versus $H_1 : p_1 - p_2 < 0$

B. $H_0 : p_1 - p_2 = 0$ versus $H_1 : p_1 - p_2 > 0$

C. $H_0 : p_1 - p_2 = 0$ versus $H_1 : p_1 - p_2 \neq 0$

iii. *Test.*

Test statistic of $\hat{p}_1 - \hat{p}_2$, with *pooled* proportion,

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{358 + 6786}{407 + 7363} =$$

(circle one) **0.9194** / **0.9934** / **0.9993**, is

$$z_0 = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}} = \frac{\left(\frac{358}{407} - \frac{6786}{7363}\right) - 0}{\sqrt{\frac{0.9194(1-0.9194)}{407} + \frac{0.9194(1-0.9194)}{7363}}} =$$

(circle one) **−1.23** / **−3.03** / **−4.56**,
so chance $\hat{p}_1 - \hat{p}_2 = \frac{358}{407} - \frac{6786}{7363} = -0.042$ or *less*, *if* $p_1 - p_2 = 0$, is p-value

$$P(\hat{p}_1 - \hat{p}_2 \leq -0.042) = P\left(\frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}} \leq -3.03\right) \approx P\left(Z \leq -3.03\right) \approx$$

(choose closest one) **0.001** / **0.025** / **0.109**.

(Stat, Proportions, Two samples, with summary, Sample 1: Number of successes: 358, Number of observations: 407, Sample 2: Number of successes: 6786, Number of observations: 7363, Next, Null: prop. = 0 Alternative: < Calculate.)

Level of significance α = (choose one) **0.01** / **0.05** / **0.10**.

iv. *Conclusion.*

Since p–value = $0.001 < \alpha = 0.050$,
do not reject / **reject** null guess: $H_0 : p_1 - p_2 = 0$.
Sample $\hat{p}_1 - \hat{p}_2$ indicates population $p_1 - p_2$
is less than / **equals** / **is greater than** 0: $H_1 : p_1 - p_2 < 0$.
In other words, the population proportion of male military doctors
is less than / **equals** / **is greater than** / **is different from**
the population proportion of male civilian doctors.

v. *Review: Population, sample, statistic and parameter.* Match columns.

terms	doctor example
(a) population	**(a)** male doctor or not, 7363 civilian, 407 military doctors
(b) sample	**(b)** $\hat{p}_1 - \hat{p}_2$
(c) statistic	**(c)** male doctor or not, all civilian/military doctors
(d) parameter	**(d)** $p_1 - p_2$

terms	(a)	(b)	(c)	(d)
doctor example				

(b) *Confidence interval.*
From above, $\hat{p}_1 = \frac{358}{407}$, $\hat{p}_2 = \frac{6786}{7363}$
and critical value for 95% $= (1-\alpha) \cdot 100\% = (1 - 0.05) \cdot 100\%$ CI,
of $z_{\frac{\alpha}{2}} = z_{\frac{0.05}{2}} = z_{0.025} \approx$ (circle one) **1.65** / **1.96** / **2.09**,
(Stat, Calculators, Normal, Mean: 0, Std. Dev.: 1, Prob(X ≥ [?]) = 0.025 Calculate.)
and so 95% CI for $p_1 - p_2$ is

$$\hat{p}_1 - \hat{p}_2 \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = \left(\frac{358}{407} - \frac{6786}{7363}\right) \pm 1.96 \cdot \sqrt{\frac{\frac{358}{407}\left(1-\frac{358}{407}\right)}{407} + \frac{\frac{6786}{7363}\left(1-\frac{6786}{7363}\right)}{7363}} =$$

(−0.054, −0.008) / **(−0.064, −0.009)** / **(−0.074, −0.010)**
(Options, choose Confidence Interval 0.95 Calculate.)

Since confidence interval does *not* include zero, this indicates population proportion of male military doctors
is less than / **equals** / **is greater than** / **is different from**
the population proportion of male civilian doctors.

3. *Inference $p_1 - p_2$, large independent samples: smokers.*
Consider number of male smokers and female smokers. Test claim there is a *different* proportion of male smokers (1) than female smokers (2) at $\alpha = 0.05$, where $x_1 = 358$, $n_1 = 958$, $x_2 = 267$ and $n_2 = 869$. Calculate 95% CI of difference in proportions.

   (a) *Hypothesis test.*

      i. *Check assumptions.*
      Since $n_1\hat{p}_1(1-\hat{p}_1) = 958 \cdot \frac{358}{958}\left(1 - \frac{358}{958}\right) \approx 224.2 > 10$, and
      $n_2\hat{p}_2(1-\hat{p}_2) = 869 \cdot \frac{267}{869}\left(1 - \frac{267}{869}\right) \approx 185.0 > 10$,
      assumptions **are** / **are not** satisfied, so continue.

      ii. *Statement.* Choose one.
         A. $H_0 : p_1 - p_2 = 0$ versus $H_1 : p_1 - p_2 < 0$
         B. $H_0 : p_1 - p_2 = 0$ versus $H_1 : p_1 - p_2 > 0$
         C. $H_0 : p_1 - p_2 = 0$ versus $H_1 : p_1 - p_2 \neq 0$

      iii. *Test.*
      Test statistic of $\hat{p}_1 - \hat{p}_2$, with *pooled* proportion,

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{358 + 267}{958 + 869} =$$

(circle one) **0.2344 / 0.3421 / 0.5993**, is

$$z_0 = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}} = \frac{\left(\frac{358}{958} - \frac{267}{869}\right) - 0}{\sqrt{\frac{0.3421(1-0.3421)}{958} + \frac{0.3421(1-0.3421)}{869}}} =$$

(circle one) **1.23 / 2.99 / 3.56**,
because of symmetry, two-sided p-value for $\hat{p}_1 - \hat{p}_2 = \frac{358}{958} - \frac{267}{869} \approx 0.066$

$$2 \cdot P(\hat{p}_1 - \hat{p}_2 \geq 0.066) = 2 \cdot P\left(\frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}} \geq 2.99\right) \approx 2 \cdot P(Z \geq 2.99) \approx$$

(choose closest one) **0.001 / 0.003 / 0.009**.

(Stat, Proportions, Two samples, with summary, Sample 1: Number of successes: 358, Number of observations: 958, Sample 2: Number of successes: 267, Number of observations: 869, Next, Null: prop. = 0 Alternative: ≠ Calculate.)

Level of significance $\alpha =$ (choose one) **0.01 / 0.05 / 0.10**.

iv. *Conclusion.*
Since p–value $= 0.003 < \alpha = 0.050$,
do not reject / reject null guess: $H_0 : p_1 - p_2 = 0$.
Sample $\hat{p}_1 - \hat{p}_2$ indicates population $p_1 - p_2$
is less than / equals / does not equal 0: $H_1 : p_1 - p_2 \neq 0$.
In other words, population proportion of male smokers
is less than / equals / is greater than / is different from
population proportion of female smokers.

(b) *Confidence interval.*
From above, $\hat{p}_1 = \frac{358}{958}$, $\hat{p}_2 = \frac{267}{869}$
and critical value for $95\% = (1 - \alpha) \cdot 100\% = (1 - 0.05) \cdot 100\%$ CI,
of $z_{\frac{\alpha}{2}} = z_{\frac{0.05}{2}} = z_{0.025} \approx$ (circle one) **1.65 / 1.96 / 2.09**,

(Stat, Calculators, Normal, Mean: 0, Std. Dev.: 1, Prob(X ≥ [?]) = 0.025 Calculate.)

and so 95% CI for $p_1 - p_2$ is

$$\hat{p}_1 - \hat{p}_2 \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = \left(\frac{358}{958} - \frac{267}{869}\right) \pm 1.96 \cdot \sqrt{\frac{\frac{358}{958}\left(1 - \frac{358}{958}\right)}{958} + \frac{\frac{267}{869}\left(1 - \frac{267}{869}\right)}{869}} =$$

(0.023, 0.110) / (0.033, 0.100) / (0.043, 0.090)

(Options, choose Confidence Interval 0.95 Calculate.)

Since confidence interval does *not* include zero, this indicates
population proportion of male smokers

is less than / equals / is greater than / is different from population proportion of female smokers.

4. *Sample size difference in proportions, given margin of error, level of confidence.* Sample size necessary to achieve a required margin of error, E, with a given level of confidence in a confidence interval for difference in proportions determined using formula, if *priors* $\hat{p}_1, \hat{p}_2$ available,

$$n = n_1 = n_2 = [\hat{p}_1(1-\hat{p}_1) + \hat{p}_2(1-\hat{p}_2)]\left(\frac{z_{\frac{\alpha}{2}}}{E}\right)^2,$$

and if priors unavailable,

$$n = n_1 = n_2 = \frac{1}{2}\left(\frac{z_{\frac{\alpha}{2}}}{E}\right)^2.$$

(a) *With priors: difference in proportion of smokers.*
Previous study gives $\hat{p}_1 = 0.35, \hat{p}_2 = 0.32$ proportions of male and female smokers, respectively. What sample sizes, n_1, n_2, required to estimate difference in proportions, $p_1 - p_2$, to within margin of error of $E = 0.01$ with 85% confidence?

$$\begin{aligned} n \;=\; n_1 \;=\; n_2 \;&=\; [\hat{p}_1(1-\hat{p}_1) + \hat{p}_2(1-\hat{p}_2)]\left(\frac{z_{\frac{\alpha}{2}}}{E}\right)^2 \\ &=\; [0.35(1-0.35) + 0.32(1-0.32)]\left(\frac{1.44}{0.01}\right)^2 \approx \end{aligned}$$

(circle one) **8888** / **9184** / **9230**.

(Since 85% confidence implies $z_{\frac{\alpha}{2}} = z_{\frac{0.15}{2}} = z_{0.075}$,
Stat, Calculators, Normal, Mean: 0, Std. Dev.: 1, Prob(X $\geq$ $\boxed{?}$) = 0.075 Calculate.)

(b) *Without priors: difference in proportion of smokers.*
Same as before but *not* knowing previous study results.

$$n_1 = n_2 = \frac{1}{2}\left(\frac{z_{\frac{\alpha}{2}}}{E}\right)^2 = \frac{1}{2}\left(\frac{1.44}{0.01}\right)^2 \approx$$

(circle one) **8409** / **9184** / **10368**.
Without priors, sample size (choose one)
decreases / **remains same** / **increases** from $n \approx 9230$ to $n \approx 10368$.

11.2 Inference about Two Means: Dependent Samples

We look at the test and confidence interval for difference in two means, dependent samples, μ_d,

$$t_0 = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}}, \quad \bar{d} \pm t_{\frac{\alpha}{2}}\left(\frac{s_d}{\sqrt{n}}\right),$$

used when data is matched pairs, and either underlying distribution of *differences* is normal with no outliers or if simple random sample size large ($n \geq 30$).

Exercise 11.2 (Inference about Two Means: Dependent Samples)

1. *Inference for difference in dependent means, μ_d: cellular response.*
A study is conducted to determine cellular response to progesterone in females. Blood cells from female 1 are broken into two groups. One group of these blood cells are injected with progesterone; the other group, the *control*, is, for comparison purposes, left untreated. Blood cells of females 2, 3 and 4 are handled in same way. Assume normality with no outliers. Test if mean progesterone response *greater* than mean control response at 5%. Calculate 95% CI.

female	progesterone (1)	control (2)
1	5.85	5.23
2	2.28	1.21
3	1.51	1.40
4	2.12	1.38

(Relabel var1 progesterone, var2 control. Type data into these two columns. Data, Data expression, Expression: progesterone - control, New column name: difference, Compute.)

(a) *Hypothesis test.*

i. *Statement.*
If mean progesterone response, μ_1, is *greater* than average control response, μ_2, $\mu_1 > \mu_2$, difference in responses must be greater than *zero*, $\mu_d = \mu_1 - \mu_2 > 0$, so (circle one)

A. $H_0 : \mu_d = 0$ versus $H_1 : \mu_d > 0$

B. $H_0 : \mu_d = 0$ versus $H_1 : \mu_d < 0$

C. $H_0 : \mu_d = 0$ versus $H_1 : \mu_d \neq 0$

ii. *Test.*

female	progesterone (1)	control (2)	differences, d_i
1	5.85	5.23	$d_1 = 5.85 - 5.23 = 0.62$
2	2.28	1.21	1.07
3	1.51	1.40	0.11
4	2.12	1.38	0.74

Average of differences is
$\bar{d} = \frac{0.62+1.07+0.11+0.74}{4}$ = (circle one) **0.355** / **0.431** / **0.635**
standard deviation of differences,
$s_d \approx$ (circle one) **0.398** / **0.931** / **1.522**,

(Stat, Summary Stats, Columns, difference, Calculate. Notice Std Dev. = 0.398.)
so test statistic of $\bar{d} = 0.635$ is

$$t_0 = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}} = \frac{0.635 - 0}{0.398/\sqrt{4}} =$$

(circle one) **1.42** / **2.93** / **3.19**,
with $n - 1 = 4 - 1 =$ **1** / **2** / **3** degrees of freedom,
so chance observed $\bar{d} = 0.635$ or *more*, *if* $\mu_d = 0$, is

$$\text{p–value} = P(\bar{d} \geq 0.635) = P\left(\frac{\bar{d} - \mu_0}{\frac{s}{\sqrt{n}}} \geq \frac{0.635 - 0}{\frac{0.398}{\sqrt{4}}}\right) \approx P\left(t \geq 3.19\right) \approx$$

(choose closest one) **0.013** / **0.025** / **0.075**.
(Stat, T statistics, Paired, Sample 1 in: progesterone, Sample 2 in: control, check Save differences, Next, Null: prop. = 0 Alternative: > Calculate.)
Level of significance $\alpha =$ (choose one) **0.01** / **0.05** / **0.10**.

iii. *Conclusion.*
Since p–value $= 0.025 < \alpha = 0.050$,
(circle one) **do not reject** / **reject** null guess: $H_0 : \mu_d = 0$.
Sample average difference $\bar{d}$ indicates population average difference μ_d (circle one) **is less than** / **equals** / **is greater than** 0: $H_1 : \mu_d > 0$.
In other words, progesterone population mean cellular response
is less than / **equals** / **is greater than** / **is different from**
control population mean cellular response.

iv. *Comment: dependent samples*
Control blood samples **depend on** / **are independent of** progesterone–infected blood samples. These dependent samples are *paired* within females. In general, *sampling is dependent if individuals in one sample are used to determine individuals in other sample.*

(b) *Confidence interval.*
Since $\bar{d} =$ (circle one) **0.355** / **0.431** / **0.635**
and $s_d \approx$ (circle one) **0.398** / **0.931** / **1.522**,
with $n - 1 = 4 - 1 =$ **1** / **2** / **3** degrees of freedom,
and critical value $95\% = (1 - \alpha) \cdot 100\% = (1 - 0.05) \cdot 100\%$ CI,
so $t_{\frac{\alpha}{2}} = t_{\frac{0.05}{2}} = t_{0.025} \approx$ (circle one) **2.28** / **2.73** / **3.19**,
(Since 95% confidence implies $z_{\frac{\alpha}{2}} = z_{\frac{0.05}{2}} = z_{0.025}$,
Stat, Calculators, T, DF: 3, Prob(X ≥ $\boxed{?}$) = 0.025 Calculate.)
and so 95% CI for μ_d is

$$\bar{d} \pm t_{\frac{\alpha}{2}} \frac{s_d}{\sqrt{n}} = 0.635 \pm 3.19 \times \frac{0.398}{\sqrt{4}} =$$

(0.0011, 1.2689) / **(0.0009, 1.2699)** / **(0.0007, 1.2709)**
(Options, Edit, choose Confidence Interval 0.95, Calculate.)

Comment: Means same or different if CIs include zero or not.
If difference in averages (close to) *zero*, averages are **same** / **different**; otherwise, averages different. Since confidence interval does *not* include zero, this indicates progesterone population mean cellular response **is less than** / **equals** / **is greater than** / **is different from** control population mean cellular response.

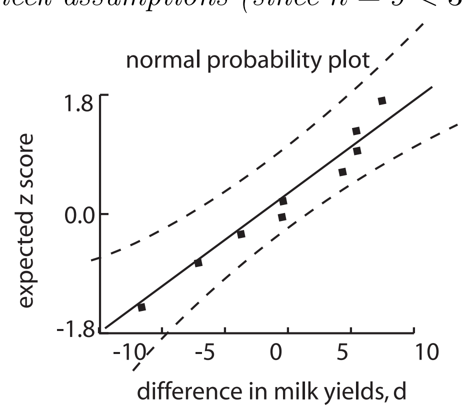
2. *Inference for difference in dependent means, μ_d: milk yield.*
A study is conducted to determine effect of "gentech" animal feed on milk yield of 9 cows. Cow 1 is fed a control feed for three months and then gentech feed for next three months for comparison purposes. Other cows are treated in same way. Test if average gentech milk yield is *greater* than average control milk yield at 5%. Calculate 95% CI.

cow	gentech (1)	control (2)	differences, d_i
1	62	54	______
2	45	43	______
3	53	55	______
4	35	39	______
5	71	65	______
6	64	62	______
7	63	56	______
8	57	50	______
9	43	52	______

(Relabel var1 gentech, var2 control. Type data into these two columns. Data, Data expression, Expression: gentech - control, New column name: difference, Compute. Data)

(a) *Hypothesis test.*

i. *Check assumptions (since $n = 9 < 30$): normality and outliers.*



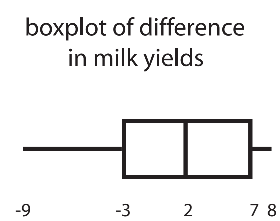


Figure 11.1 (Normal probability plot, boxplot for milk yield differences.)

A. *Data normal?*
Normal probability plot indicates differences
normal / **not normal** because data *outside* dotted bounds
(Notice dotted lines flare out at lower end, so includes points.)
Graphics, QQ Plot, Select Columns: differences, Create Graph!

B. *Outliers?*
Boxplot indicates **outliers** / **no outliers**.
Graphics, Boxplot, select differences, Next, check Use fences to identify outliers, check Draw boxes horizontally. Create Graph!

ii. *Statement.*
A. $H_0 : \mu_d = 0$ versus $H_1 : \mu_d < 0$
B. $H_0 : \mu_d = 0$ versus $H_1 : \mu_d > 0$
C. $H_0 : \mu_d = 0$ versus $H_1 : \mu_d \neq 0$

iii. *Test.*
$\bar{d} \approx$ (circle one) **1.41** / **1.89** / **2.52**,
$s_d \approx$ (circle one) **5.47** / **5.86** / **6.52**,
(Stat, Summary Stats, Columns, difference, Calculate.)
so test statistic of $\bar{d} = 1.89$ is

$$t_0 = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}} = \frac{1.89 - 0}{5.86/\sqrt{9}} =$$

(circle one) **0.42** / **0.97** / **1.34**,
with $n - 1 = 9 - 1 =$ **6** / **7** / **8** degrees of freedom,
so chance observed $\bar{d} = 1.89$ or *more*, *if* $\mu_d = 0$, is

$$\text{p–value} = P(\bar{d} \geq 1.89) = P\left(\frac{\bar{d} - \mu}{\frac{s}{\sqrt{n}}} \geq \frac{1.89 - 0}{5.86/\sqrt{9}}\right) \approx P\left(t \geq 0.97\right) \approx$$

(circle one) **0.12** / **0.15** / **0.18**.
(Stat, T statistics, Paired, Sample 1 in: gentech, Sample 2 in: control, check Save differences, Next, Null: prop. = 0 Alternative: > Calculate.)
Level of significance $\alpha =$ (choose one) **0.01** / **0.05** / **0.10**.

iv. *Conclusion.*
Since p–value $= 0.18 > \alpha = 0.05$,
(circle one) **do not reject** / **reject** null guess: $H_0 : \mu_d = 0$.
Sample average difference $\bar{d}$ indicates population average difference μ_d
(circle one) **is less than** / **equals** / **is greater than** 0: $H_0 : \mu_d = 0$.
In other words, gentech population mean milk yield
is less than / **equals** / **is greater than** / **is different from**
control population mean milk yield.

(b) *Confidence interval.*
$\bar{d} \approx$ (circle one) **1.41 / 1.89 / 2.52**,
$s_d \approx$ (circle one) **5.47 / 5.86 / 6.52**,
with $n - 1 = 9 - 1 =$ **6 / 7 / 8** degrees of freedom,
and critical value $95\% = (1 - \alpha) \cdot 100\% = (1 - 0.05) \cdot 100\%$ CI,
so $t_{\frac{\alpha}{2}} = t_{\frac{0.05}{2}} = t_{0.025} \approx$ (circle one) **2.31 / 2.53 / 3.09**,
(Since 95% confidence implies $z_{\frac{\alpha}{2}} = z_{\frac{0.05}{2}} = z_{0.025}$,
Stat, Calculators, T, DF: 8, Prob(X $\geq$ [?]) = 0.025 Calculate.)
and so 95% CI for μ_d is

$$\bar{d} \pm t_{\frac{\alpha}{2}} \frac{s_d}{\sqrt{n}} = 1.89 \pm 2.31 \times \frac{5.86}{\sqrt{8}} =$$

(−2.52, 6.49) / (−2.62, 6.39) / (−2.72, 6.29)
(Options, Edit, choose Confidence Interval 0.95, Calculate.)

Since confidence interval *does* include zero, this indicates gentech population mean milk yield
is less than / equals / is greater than / is different from
control population mean milk yield.

11.3 Inference about Two Means: Independent Samples

We look at the test and confidence interval for difference in two means, *in*dependent samples, $\mu_1 - \mu_2$,

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \qquad (\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}},$$

used when data is independent, and either underlying distribution of both samples are normal with no outliers or if both simple random sample sizes large ($n_1 \geq 30, n_2 \geq 30$) and where degrees of freedom for test is

$$\text{df} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}.$$

This is called the *Satterthwaite's formula* for degrees of freedom. StatCrunch uses this formula to determine the degrees of freedom. Although text suggests approximating degrees of freedom by setting it equal to minimum of $n_1 - 1$ and $n_2 - 1$, we will always use exact Satterthwaite formula to calculate degrees of freedom in this case.

Exercise 11.3 (Inference about Two Means: Independent Samples)

1. *Inference for $\mu_1 - \mu_2$, independent samples, unknown σ, raw data: progesterone.* A study is conducted to determine cellular response to progesterone in females. Blood cells from four females are injected with progesterone; blood cells from four *different* females are, for comparison purposes, left untreated. Test if average progesterone response is *greater* than average control response at 5%. Calculate 95% CI. Assume normality with no outliers.

female	progesterone (1)	female	control (2)
1	5.85	5	5.23
2	2.28	6	1.21
3	1.51	7	1.40
4	2.12	8	1.38

(Blank data table. StatCrunch, My Data.)

(a) *Hypothesis test.*

i. *Statement.* Choose one.

A. $H_0 : \mu_1 - \mu_2 = 0$ versus $H_1 : \mu_1 - \mu_2 < 0$
B. $H_0 : \mu_1 - \mu_2 = 0$ versus $H_1 : \mu_1 - \mu_2 > 0$
C. $H_0 : \mu_1 - \mu_2 = 0$ versus $H_1 : \mu_1 - \mu_2 \neq 0$

ii. *Test.*

female	progesterone (1)	female	control (2)
1	5.85	5	5.23
2	2.28	6	1.21
3	1.51	7	1.40
4	2.12	8	1.38
average	$\bar{x}_1 = \frac{5.85+2.28+1.51+2.12}{4} = 2.94$		$\bar{x}_2 = 2.305$
SD	$s_1 = 1.97$		$s_2 = 1.95$

Test statistic of $\bar{x}_1 - \bar{x}_2 = 0.635$ is

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(2.94 - 2.305) - 0}{\sqrt{\frac{1.97^2}{4} + \frac{1.95^2}{4}}} =$$

(circle one) **0.458** / **2.93** / **4.56**,
with degrees of freedom

$$\text{df} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{s_2^2}{n_2}\right)^2} = \frac{\left(\frac{1.97^2}{4} + \frac{1.95^2}{4}\right)^2}{\frac{1}{4-1}\left(\frac{1.97^2}{4}\right)^2 + \frac{1}{4-1}\left(\frac{1.95^2}{4}\right)^2} = 5.999$$

so chance observed $\bar{x}_1 - \bar{x}_2 = 0.635$ or more, *if* $\mu_1 - \mu_2 = 0$, is p-value

$$P(\bar{x}_1 - \bar{x}_2 \geq 0.635) = P\left(\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \geq \frac{(2.94 - 2.305) - 0}{\sqrt{\frac{1.97^2}{4} + \frac{1.95^2}{4}}}\right) \approx P\left(t \geq 0.458\right)$$

equals (choose closest one) **0.13** / **0.25** / **0.33**.

(Stat, T statistics, Two sample, with data, Sample 1 in: progesterone, Sample 2 in: control, do *not* check Pool variances, Next, Null: prop. = 0 Alternative: > Calculate.)

Level of significance $\alpha =$ (choose one) **0.01** / **0.05** / **0.10**.

iii. *Conclusion.*

Since p–value = $0.33 > \alpha = 0.050$,
do not reject / **reject** null guess: $H_0 : \mu_1 - \mu_2 = 0$.
Sample average difference $\bar{x}_1 - \bar{x}_2$ indicates population difference $\mu_1 - \mu_2$ **is less than** / **equals** / **is greater than** 0: $H_0 : \mu_1 - \mu_2 = 0$.
In other words, progesterone population mean cellular response **is less than** / **equals** / **is greater than** / **is different from** control population mean cellular response.

iv. *Comment: independent samples*

Control blood samples for four women **depend on** / **are independent of** progesterone–infected blood samples of four other women. In general, *sampling is independent if individuals in one sample do not determine individuals in other sample.* Present test, here, in independent case probably a *worse* test than previous test for same data in dependent case. Independent test does not control for differences due to different females, whereas previous test does do this.

v. *Population, Sample, Statistic, Parameter.* Match columns.

terms	infection example
(a) populations	**(a)** cellular response, 8 females
(b) samples	**(b)** $\bar{x}_1 - \bar{x}_2$
(c) statistic	**(c)** cellular response, all females
(d) parameter	**(d)** $\mu_1 - \mu_2$

terms	(a)	(b)	(c)	(d)
infection example				

(b) *Confidence interval.*

From above, $\bar{x}_1 \approx 2.94$, $s_1 \approx 1.97$, $\bar{x}_2 \approx 2.305$, $s_2 \approx 1.95$, df ≈ 5.999, and critical value for 95% = $(1 - \alpha) \cdot 100\% = (1 - 0.05) \cdot 100\%$ CI, of $t_{\frac{\alpha}{2}} = t_{\frac{0.05}{2}} = t_{0.025} \approx$ (circle one) **2.31** / **2.45** / **3.09**,

(Since 95% confidence implies $t_{\frac{\alpha}{2}} = t_{\frac{0.05}{2}} = t_{0.025}$,
Stat, Calculators, T, DF: 6, Prob(X ≥ [?]) = 0.025 Calculate.)

and so 95% CI for $\mu_1 - \mu_2$ is

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = (2.94 - 2.305) \pm 2.45 \cdot \sqrt{\frac{1.97^2}{4} + \frac{1.95^2}{4}} =$$

(−2.52, 6.49) / **(−2.62, 6.39)** / **(−2.76, 4.03)**

(Options, Edit, choose Confidence Interval 0.95, Calculate.)

Since confidence interval *does* include zero, this indicates
progesterone population mean cellular response
is less than / equals / is greater than / is different from
control population mean cellular response.

2. *Inference for differences in means: plasma levels.*
Consider statistics on plasma levels for a random sample of nine 17–year–old males and six 17–year–old females. Test if two means different, $\mu_1 \neq \mu_2$, at 5%. Calculate 95% CI. Assume normality with no outliers.

	males (1)	females (2)
$\bar{x}$	3.259	1.413
s	0.400	0.220
n	9	6

Notice this is a *summary* of data, with $\bar{x}$, s and n, *not* a list of data!

(a) *Hypothesis test.*

i. *Statement.* Choose one.
A. $H_0: \mu_1 - \mu_2 = 0$ versus $H_1: \mu_1 - \mu_2 < 0$
B. $H_0: \mu_1 - \mu_2 = 0$ versus $H_1: \mu_1 - \mu_2 > 0$
C. $H_0: \mu_1 - \mu_2 = 0$ versus $H_1: \mu_1 - \mu_2 \neq 0$

ii. *Test.*
Test statistic of $\bar{x}_1 - \bar{x}_2 = 3.259 - 1.413 = 1.846$ is

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(3.259 - 1.413) - 0}{\sqrt{\frac{0.4^2}{9} + \frac{.22^2}{6}}} =$$

(circle one) **9.45** / **10.23** / **11.48**,
with degrees of freedom

$$\text{df} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{s_2^2}{n_2}\right)^2} = \frac{\left(\frac{0.4^2}{9} + \frac{0.22^2}{6}\right)^2}{\frac{1}{9-1}\left(\frac{0.4^2}{9}\right)^2 + \frac{1}{6-1}\left(\frac{0.22^2}{6}\right)^2} = 12.72$$

so chance $\bar{x}_1 - \bar{x}_2 = 3.259 - 1.413 = 1.846$ or more, *if* $\mu_1 - \mu_2 = 0$, is p-value

$$P(\bar{x}_1 - \bar{x}_2 \geq 1.846) = P\left(\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \geq \frac{(3.259 - 1.413) - 0}{\sqrt{\frac{0.4^2}{9} + \frac{.22^2}{6}}}\right) \approx P\left(t \geq 11.48\right)$$

equals (choose closest one) **0.00** / **0.05** / **4.42**.

(Stat, T statistics, Two sample, with summary, Sample 1 Mean: 3.259, Standard deviation: 0.4, Size: 9, Sample 2 Mean: 1.413, Standard deviation: 0.22, Size: 6, do *not* check Pool variances, Next, Null: prop. = 0 Alternative: $\neq$ Calculate.)

Level of significance $\alpha =$ (choose one) **0.01** / **0.05** / **0.10**.

iii. *Conclusion.*

Since p–value $= 0.000 < \alpha = 0.050$,
do not reject / **reject** null guess: $H_0 : \mu_1 - \mu_2 = 0$.
Sample $\bar{x}_1 - \bar{x}_2$ indicates population $\mu_1 - \mu_2$
is less than / **equals** / **does not equal** 0: $H_0 : \mu_1 - \mu_2 \neq 0$.
In other words, average male plasma amount
is less than / **equals** / **is greater than** / **is different from**
average female plasma amount.

iv. *Population, Sample, Statistic, Parameter.* Match columns.

terms	plasma example
(a) population	**(a)** nine male, six female plasma levels
(b) sample	**(b)** $\bar{x}_1 - \bar{x}_2$
(c) statistic	**(c)** all 17–year–olds plasma levels
(d) parameter	**(d)** $\mu_1 - \mu_2$

terms	(a)	(b)	(c)	(d)
plasma example				

(b) *Confidence interval.*

From above, $\bar{x}_1 = 3.259$, $s_1 = 0.4$, $\bar{x}_2 = 1.413$, $s_2 = 0.22$, df ≈ 12.72, and critical value for 95% $= (1 - \alpha) \cdot 100\% = (1 - 0.05) \cdot 100\%$ CI, of $t_{\frac{\alpha}{2}} = t_{\frac{0.05}{2}} = t_{0.025} \approx$ (circle one) **2.17** / **2.45** / **3.09**,

(Since 95% confidence implies $t_{\frac{\alpha}{2}} = t_{\frac{0.05}{2}} = t_{0.025}$,
Stat, Calculators, T, DF: 12.72, Prob(X $\geq$ $\boxed{?}$) = 0.025).

and so 95% CI for $\mu_1 - \mu_2$ is

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = (3.259 - 1.413) \pm 2.17 \cdot \sqrt{\frac{0.4^2}{4} + \frac{0.22^2}{4}} =$$

(1.50, 2.19) / **(1.55, 2.09)** / **(1.60, 1.99)**

(Options, Edit, choose Confidence Interval 0.95, Calculate.)

Since confidence interval does *not* include zero, this indicates
average male plasma amount
is less than / **equals** / **is greater than** / **is different from**
average female plasma amount.

11.4 Inference about Two Population Standard Deviations

After look at F distribution, then use this distribution to test ratio of two variances with test statistic

$$F_0 = \frac{s_1^2}{s_2^2},$$

which can be used when both underlying distributions normal with no outliers and simple random samples collected.

Exercise 11.4 (Inference about Two Population Standard Deviations)

1. *Test* $\frac{\sigma_1}{\sigma_2}$*: plasma levels.*
   Consider statistics on plasma levels for males and females given in table. Test if $\sigma_1 > \sigma_2$ at $\alpha = 0.05$. Assume both simple random samples normal with no outliers and collected independently of one another.

	males (1)	females (2)
$\bar{x}$	3.259	1.413
s	0.16	0.09
n	9	6

   (a) *Statement.* Choose one.

   i. $H_0 : \sigma_1 = \sigma_2$ versus $H_1 : \sigma_1 > \sigma_2$
   ii. $H_0 : \sigma_1 = \sigma_2$ versus $H_1 : \sigma_1 < \sigma_2$
   iii. $H_0 : \sigma_1 = \sigma_2$ versus $H_1 : \sigma_1 \neq \sigma_2$

   Since $\sigma_1 = \sigma_2$, $\frac{\sigma_1}{\sigma_2} = 1$,
   so $H_0 : \sigma_1 = \sigma_2$ is the same as $H_0 : \frac{\sigma_1}{\sigma_2} = 1$.

   (b) *Test.*
   With $n_1 - 1 = 9 - 1 = 8$ df and $n_2 - 1 = 6 - 1 = 5$ df,
   chance observed $\frac{s_1^2}{s_2^2} = \frac{0.16^2}{0.09^2} \approx 3.16$ or *more*, *if* $\sigma_1 = \sigma_2$, is

$$\text{p–value} = P\left(F \geq \frac{s_1^2}{s_2^2}\right) = P\left(F \geq \frac{0.16^2}{0.09^2}\right) \approx P\left(F \geq 3.16\right) \approx$$

   0.02 / **0.04** / **0.11**.
   (Stat, Variance, Two sample, with summary, Sample 1 Variance: 0.0256 (which is 0.16^2), Size: 9, Sample 2 Variance: 0.0081 (which is 0.09^2), Size: 6, Next, Null: variance ratio = 1 Alternative: > Calculate.)
   Level of significance $\alpha =$ (choose one) **0.01** / **0.05** / **0.11**.

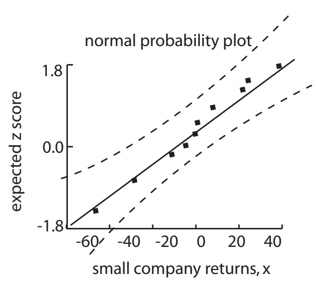
(c) *Conclusion.*
Since p–value $= 0.11 > \alpha = 0.05$,
do not reject / reject null guess: $H_0 : \sigma_1 = \sigma_2$.
Sample $\frac{s_1^2}{s_2^2}$ indicates population $\frac{\sigma_1}{\sigma_2}$
is greater than / equals / does not equal 1: $H_0 : \sigma_1 = \sigma_2$.
In other words, population SD in male plasma
is less than / equals / is greater than / is different from
population proportion SD in female plasma.

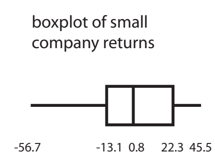
2. *Test of* $\frac{\sigma_1}{\sigma_2}$, $s_1 \neq s_2$, *raw data: company returns.*
Consider simple random sample of returns for small (1) and large (2) companies, collected independently of one another. Test $\sigma_1 \neq \sigma_2$ at $\alpha = 0.05$.

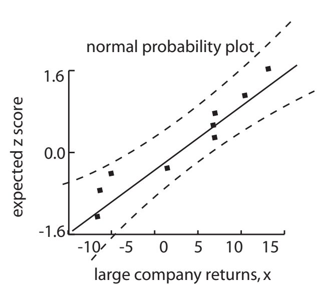
company	return	company	return
small	1.3	large	9.1
small	22.3	small	0.3
small	-30.3	large	9.4
large	15.4	small	-13.3
small	23.1	small	-3.5
small	-56.7	small	45.4
large	3.4	small	10.3
large	-7.5	large	11.9
large	-8.8	large	-9.2
		large	9

Blank data table. Relabel var1 company, var2 return, Type data into these two columns.

(a) *Check assumptions (since* $n_1 = 10 < 30, n_2 = 10 < 30$*).*







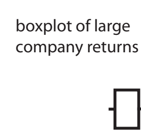


Figure 11.2 (Normal probability plots, boxplots for company returns.)

i. *Data normal?*
Normal probability plots for *both* samples indicates
normal / **not normal** because data within dotted bounds.
(Notice dotted lines flare out at lower end, so includes points.)
Graphics, QQ lot, Select Columns: small company returns and large company returns, Create Graph!

ii. *Outliers?*
Both boxplots indicates **outliers** / **no outliers**.
Graphics, Boxplot, select small company returns and large company returns, Next, check Use fences to identify outliers, check Draw boxes horizontally. Create Graph!

(b) *Statement.* Choose one.

i. $H_0 : \sigma_1 = \sigma_2$ versus $H_1 : \sigma_1 > \sigma_2$
ii. $H_0 : \sigma_1 = \sigma_2$ versus $H_1 : \sigma_1 < \sigma_2$
iii. $H_0 : \sigma_1 = \sigma_2$ versus $H_1 : \sigma_1 \neq \sigma_2$

(c) *Test (classical approach).*
Since $\alpha = 0.05$, $n_1 = 10$, $n_2 = 9$,
Lower critical value $F_{1-\frac{\alpha}{2},n_1-1,n_2-1} = F_{1-\frac{0.05}{2},10-1,9-1} = F_{0.975,9,8} =$
(choose one) **0.24** / **0.56** / **0.71 1.36**.
(Stat, Calculators, F, Num. DF: 9, Den. DF: 8, Prob(X $\geq$ $\boxed{?}$) = 0.975)
Upper critical value $F_{\frac{\alpha}{2},n_1-1,n_2-1} = F_{\frac{0.05}{2},10-1,9-1} = F_{0.025,9,8} =$
(choose one) **2.61** / **3.56** / **3.71** / **4.36**.
(Stat, Calculators, F, Num. DF: 9, Den. DF: 8,Prob(X $\geq$ $\boxed{?}$) = 0.025)
and test statistic $F_0 \approx \frac{s_1^2}{s_2^2} \approx$ (choose one) **7** / **8** / **9** / **10**.
(Stat, Variance Stats, Two Sample, With Data, Sample 1 in: return, Where: (type) company = small, Sample 2 in: return, Where: (type) company = large, $H_0 : \sigma_1^2/\sigma_2^2 = 1$, $H_A : \sigma_1^2/\sigma_2^2 \neq 1$, Calculate. Notice F-stat = 9.0107532.)

(d) *Conclusion.*
Since $F_{0.975,9,8} \approx 0.24 < F_{0.025,9,8} \approx 4.36 < F_0 \approx 9$,
(test statistic F_0 is *in* the critical region, so "away" from null)
do not reject / **reject** null guess: $H_0 : \sigma_1 = \sigma_2$.
Sample $\frac{s_1^2}{s_2^2}$ indicates population $\frac{\sigma_1}{\sigma_2}$
is less than / **equals** / **does not equal** 1: $H_1 : \sigma_1 \neq \sigma_2$.
In other words, data indicates SD in small company returns is
(circle one) **less than** / **the same as** / **different from**
in fact, probably larger than SD in large company returns.

3. *Probabilities for F distribution.*
At McDonalds in Westville, waiting time to order (in minutes) follows an F

distribution. Consider following figure with two F distributions, each with different shaded areas (probabilities).

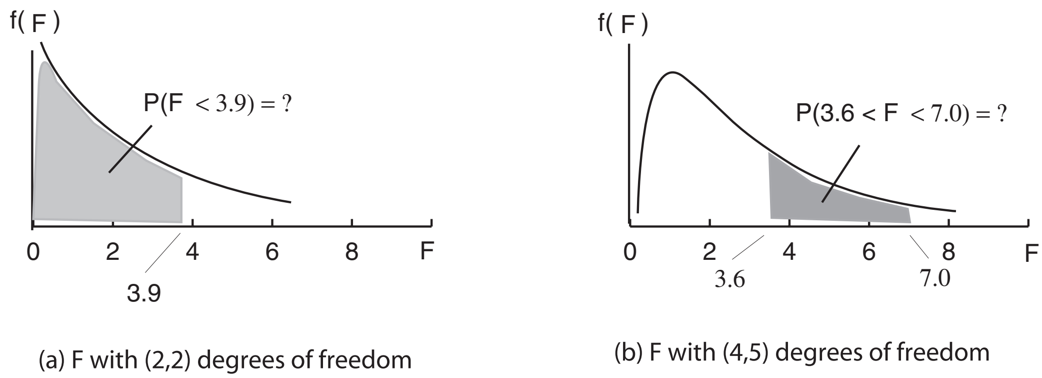


Figure 11.3 (Probabilities for F distribution.)

(a) For a F with (2,2) df, probability of waiting less than 3.9 minutes $P(F < 3.9) =$ (circle one) **0.35** / **0.45** / **0.80** / **0.92**.
(Stat, Calculators, F, Num. DF: 2, Den. DF: 2, Prob(X ≤) 3.9 = [?])

(b) For an F with (4,5) degrees of freedom, $P(3.6 < F < 7.0) =$ (circle one) **0.03** / **0.07** / **0.09** / **0.11**.
Stat, Calculators, F, Between, Num. DF: 4, Den. DF: 5, Prob(3.6 ≤ X ≤ 7.0) = [?] Compute.

(c) F distribution with (2,2) degrees of freedom in diagram (a) above is (circle one) **skewed right** / **symmetric** / **skewed left**.

(d) F distribution with (4,5) degrees of freedom in diagram (b) above is (circle one) **skewed right** / **symmetric** / **skewed left**.

(e) In general, F distribution is **symmetric** / **asymmetric**

(f) Total area (probability) under this curve is (circle one) **50%** / **75%** / **100%** / **150%**.

(g) F distribution is indexed by *two* degrees of freedom (df_1, df_2), which are equal to $df_1 = n_1 - 1$ and $df_2 = n_2 - 1$. For samples of size $n_1 = 10$ and $n_2 = 11$, F distribution has $\mathbf{(9, 10)}$ / $\mathbf{(10, 11)}$ / $\mathbf{(11, 12)}$ df.

4. *Percentiles for F distribution.*
At McDonalds in Westville, waiting time to order (in minutes) follows an F distribution. Consider following figure with two F distributions, each with 72nd percentile waiting time.

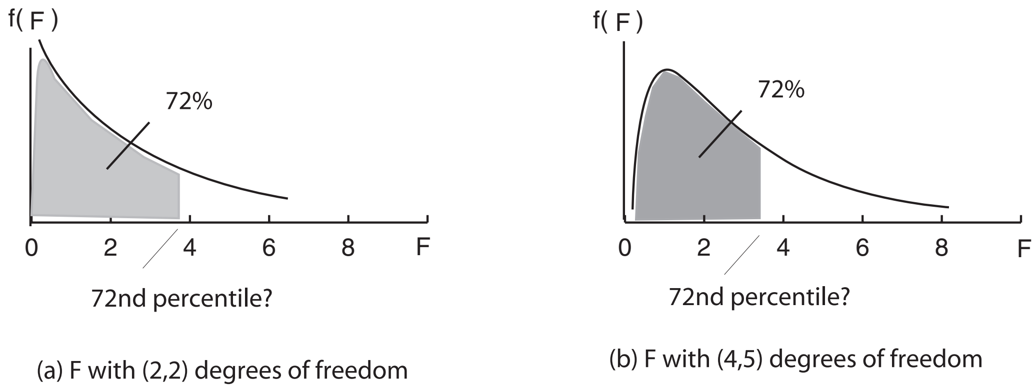


Figure 11.4 (Percentiles for F distribution.)

(a) The 72nd percentile for F with (2,2) degrees of freedom, is (circle one) **0.5** / **1.7** / **2.6** / **3.1**.
(Stat, Calculators, F, Num. DF: 2, Den. DF: 2, Prob(X $\leq$) $\boxed{?}$ = 0.72)

(b) The 72nd percentile for F with (4,5) degrees of freedom, is (circle one) **0.5** / **1.7** / **2.6** / **3.1**.

(c) If $\alpha = 0.20$, $n_1 = 19$, $n_2 = 8$
$F_{\frac{\alpha}{2},n_1-1,n_2-1} = F_{\frac{0.20}{2},19-1,8-1} = F_{0.10,18,7} =$ **2.61** / **2.56** / **2.71** / **2.80**.
(Stat, Calculators, F, Num. DF: 18, Den. DF: 7, Prob(X $\geq$) $\boxed{?}$ = 0.10)

(d) If $\alpha = 0.20$, $n_1 = 19$, $n_2 = 8$
$F_{1-\frac{\alpha}{2},n_2-1,n_1-1} = F_{1-\frac{0.20}{2},8-1,19-1} = F_{0.90,7,18} =$ **0.38** / **0.51** / **0.64** / **0.79**.
(Stat, Calculators, F, Num. DF: 7, Den. DF: 18, Prob(X $\geq$) $\boxed{?}$ = 0.90)
Notice $F_{0.10,18,7} = \frac{1}{F_{0.90,7,18}}$ and, in general,

$$F_{\frac{\alpha}{2},n_1-1,n_2-1} = \frac{1}{F_{1-\frac{\alpha}{2},n_2-1,n_1-1}}$$

11.5 Putting it Together: Which Method Do I Use?

HYPOTHESIS TESTS		mean μ	variance σ^2	proportion p
	one	chapter 10	chapter 10	chapter 10
sample	two	independent: $t_0 = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$	$F_0 = \frac{s_1^2}{s_2^2}$	independent: $z_0 = \frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1}+\frac{\hat{p}(1-\hat{p})}{n_2}}}$ where $\hat{p} = \frac{x_1+x_2}{n_1+n_2}$
		dependent: $t_0 = \frac{d-\mu_d}{s_d/\sqrt{n}}$		dependent: $z_0 = \frac{\lvert f_{12}-f_{21}\rvert-1}{\sqrt{f_{12}+f_{21}}}$
	multiple	chapter 13	not covered	chapter 12

CONFIDENCE INTERVALS		mean μ	variance σ^2	proportion p
	one	chapter 9	chapter 9	chapter 9
sample	two	independent: $(\bar{x}_1 - \bar{x}_2) \pm t_{\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	not covered	$\hat{p}_1 - \hat{p}_2 \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$
		dependent: $\bar{d} \pm t_{\frac{\alpha}{2}} \left(\frac{s_d}{\sqrt{n}}\right)$		
	multiple	not covered	not covered	not covered