Statistics 5041
11. Hotelling's T^2

Gary W. Oehlert
School of Statistics
313B Ford Hall
612-625-1557
gary@stat.umn.edu

Let's think about the univariate t-test.
First recall that there are one-sample tests, two-sample tests, paired tests, and so on. Start with the one-sample situation.
$x_1, x_2, \ldots, x_n$ are iid $\mathrm{N}(\mu, \sigma^2)$, with both μ and σ unknown. $\bar{x}$ estimates μ, and s estimates σ.
$\bar{x} \sim \mathrm{N}(\mu, \sigma^2/n)$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

or

$$t^2 = n(\bar{x} - \mu)(s^{-2})(\bar{x} - \mu) \sim \mathrm{F}_{1,n-1}$$

To test $\mathrm{H}_0 : \mu = \mu_0$ versus $\mathrm{H}_a : \mu \neq \mu_0$, reject if $|t|$ is too big or if t^2 is too big. Compute p-values by comparison with reference distributions.
We assumed normality, but we can get away from that for large sample sizes. As long as the data are iid with finite variance,

$$t \to \mathrm{N}(0,1) = t_\infty \quad \text{as} \quad n \to \infty$$

and

$$t^2 \to \chi^2_1 = \mathrm{F}_{1,\infty} \quad \text{as} \quad n \to \infty$$

We can also produce confidence intervals.
The $1 - \alpha$ confidence interval for μ is the set of potential values for μ that yield p-values of α or more in the t or t^2 test.

$$\{\mu : |t| < t_{\alpha/2,n-1}\} = \{\mu : t^2 < \mathrm{F}_{\alpha,1,n-1}\} =$$

$$(\bar{x} - t_{\alpha/2,n-1}\frac{1}{\sqrt{n}}, \ \ \bar{x} + t_{\alpha/2,n-1}\frac{1}{\sqrt{n}})$$

The *paired* setup has iid data pairs (x_i, y_i), with the assumptions that the differences $d_i = x_i - y_i$ are iid distributed $\mathrm{N}(\mu, \sigma^2)$.
Just use one-sample procedures on the differences, using $\bar{d}$ and s_d (still $n-1$ degrees of freedom).
Two-sample procedures. Assume that $x_1, x_2, \ldots, x_n$ are iid $\mathrm{N}(\mu_1, \sigma_1^2)$, and that $y_1, y_2, \ldots, y_m$ are iid $\mathrm{N}(\mu_2, \sigma_2^2)$.
Inference about $\mu_1 - \mu_2$.
If we believe $\sigma_1 = \sigma_2 = \sigma$, we can use *pooled* procedures.
If we allow $\sigma_1 \neq \sigma_2$, we must use *unpooled* procedures.
Pooling.
Let $s_p^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}$. Under $\mathrm{H}_0 : \mu_1 - \mu_2 = 0$,

$$\frac{\bar{x} - \bar{y}}{\sqrt{(1/n + 1/m)s_p^2}} \sim t_{n+m-2}$$

or

$$(1/n + 1/m)^{-1}(\bar{x} - \bar{y})s_p^{-2}(\bar{x} - \bar{y}) \sim \mathrm{F}_{1,n+m-2}$$

Confidence interval for $\mu_1 - \mu_2$:

$$\bar{x} - \bar{y} \pm t_{\alpha/2,n-1}\sqrt{1/n + 1/m}\; s_p$$

The pooled procedures work in large samples even for nonnormally distributed data, if the variances are equal.
The pooled procedures do *not* work if $\sigma_1 \neq \sigma_2$ and can give misleading results.
Unpooled procedures.

$$t_p = \frac{\bar{x} - \bar{y}}{\sqrt{s_x^2/n + s_y^2/m}}$$

is only approximately t distributed. Use t with Satterthwaite approximate degrees of freedom for small n and m.

$$df = \frac{(s_x^2/n + s_y^2/m)^2}{\frac{1}{n-1}\frac{s_x^4}{n^2} + \frac{1}{m-1}\frac{s_y^4}{m^2}}$$

t_p is approximately standard normal for large n and m.
Form confidence intervals or t^2 test in the usual way.
What do we do for multivariate data? We use *Hotelling's* T^2.
For a one-sample problem, x_i *iid* $\mathrm{N}_p(\mu, \Sigma)$, testing $\mathrm{H}_0 : \mu = \mu_0$

$$T^2 = (\overline{\mathbf{x}} - \mu_0)' \left(\frac{1}{n}\mathbf{S}\right)^{-1} (\overline{\mathbf{x}} - \mu_0) = n(\overline{\mathbf{x}} - \mu_0)'\mathbf{S}^{-1}(\overline{\mathbf{x}} - \mu_0)$$

T^2 is the squared Mahalanobis distance (with estimated variance) between the observed mean and the null hypothesis mean.
For large n, T^2 is approximately χ_p^2 under the null hypothesis.
For small n,

$$T^2 \sim \frac{(n-1)p}{(n-p)}F_{p,\,n-p}$$

under the null hypothesis.
The p-value for the test is thus

$$P(F_{p,n-p} > \frac{(n-p)}{(n-1)p}T^2)$$

To construct a $1 - \alpha$ confidence region for μ, use

$$\left\{\mu : n(\overline{\mathbf{x}} - \mu)'\mathbf{S}^{-1}(\overline{\mathbf{x}} - \mu) \leq \frac{(n-1)p}{(n-p)}F_{\alpha,\,p,\,n-p}\right\}$$

This confidence region is an ellipsoid centered at $\overline{\mathbf{x}}$ with axes oriented along the eigenvectors of $\mathbf{S}$ and axis lengths proportional to the square roots of the eigenvalues of $\mathbf{S}$.
Try wood stiffness data from text.

```
Cmd> readdata("",x1,x2,x3,x4,x5)
Read from file "/cdrom/T4-3.DAT"
Column 1 saved as REAL vector x1
```

```
Column 2 saved as REAL vector x2
Column 3 saved as REAL vector x3
Column 4 saved as REAL vector x4
Column 5 saved as REAL vector x5

Cmd> X <- hconcat(x1,x2,x3,x4)

Cmd> xbar <- tabs(X,mean:T);xbar
(1)   1906.1   1749.5   1509.1     1725

Cmd> S <- tabs(X,covar:T)
```

We have the null of all means at 1750.

```
Cmd> mu0 <- rep(1750,4)

Cmd> T2 <- (xbar - mu0)'%*%solve(S)%*%\
(xbar - mu0)*30

Cmd> T2
(1,1)       277.95

Cmd> T2*(30-4)/(30-1)/4  # F distributed
(1,1)         62.3

Cmd> 1-cumF(62.3,4,26)
(1)   6.1018e-13
```

Tiny p-value. Can we find where differences are?

```
Cmd> U <- eigen(S)$vectors

Cmd> lam <- eigenvals(S)

Cmd> (U'%*%(xbar-mu0))/sqrt(lam/30)
(1,1)     -0.41258
(2,1)      -5.0143
(3,1)      -12.831
(4,1)       9.3808

Cmd> 12.83^2+9.38^2+5.01^2+.41^2
(1)       277.86

Cmd> U
(1,1)   0.526  -0.199  -0.240   0.791
(2,1)   0.487  -0.727   0.136  -0.465
(3,1)   0.476   0.445   0.759   0.025
(4,1)   0.510   0.484  -0.590  -0.396
```

First element of `(U'%*%(xbar-mu0))/sqrt(lam/30)` was OK, but others were huge.
First column of `U` is more or less constant, corresponding to the average of the elements of `xbar-mu0`. The others are differences between elements, and they are all too big.
For ease of visualization, just do confidence region for first two variables.

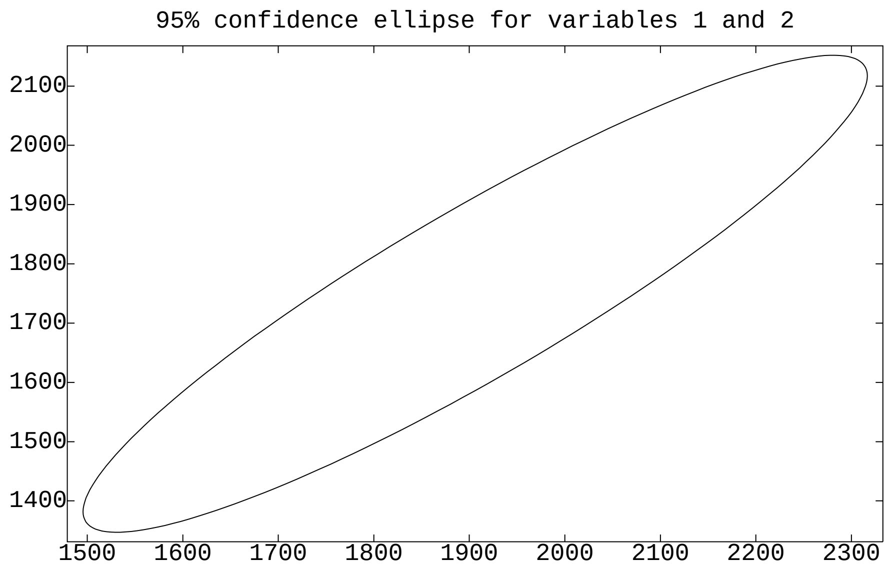
```
Cmd> xbar12 <- xbar[vector(1,2)]

Cmd> S12 <- S[vector(1,2),vector(1,2)]

Cmd> 2*(30-1)/(30-2)*invF(.95,2,28)
(1)       6.9194

Cmd> ellipse(6.919,S12/30,xbar12,draw:T)

Cmd> showplot(title:"95% confidence ellipse\
 for variables 1 and 2")
```



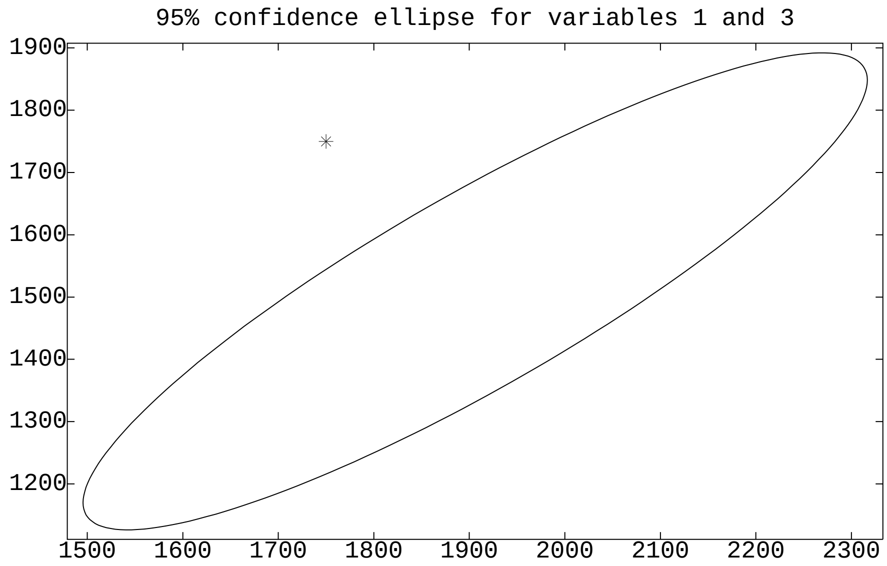

```
Cmd> xbar13 <- xbar[vector(1,3)]

Cmd> S13 <- S[vector(1,3),vector(1,3)]

Cmd> ellipse(6.919,S13/30,xbar13,draw:T)

Cmd> addpoints(1750,1750)

Cmd> showplot(title:"95% confidence ellipse\
 for variables 1 and 3")
```



Let's be a little more particular about what is happening.
Let $w \sim \mathrm{N}_p(0, a\Sigma)$ under H_0.
Let $\mathbf{V} \sim W_f(a\Sigma)$ independent of w.
Then

$$w'\mathbf{V}^{-1}w \sim \frac{fp}{f-p+1}F_{p,\, f-p+1}$$

For the one-sample T^2, $f = n-1$, $a = 1/n$.
For a multivariate paired problem, we again take differences and use one-sample T^2 with $f = n-1$ and $a = 1/n$.
For *pooled* two-sample T^2 under H_0

$$(\overline{\mathbf{x}} - \overline{\mathbf{y}}) \sim \mathrm{N}_p(0, (\frac{1}{n} + \frac{1}{m})\Sigma)$$

$$\mathbf{V} = \mathbf{S}_p = \frac{(n-1)\mathbf{S}_x + (m-1)\mathbf{S}_y}{n+m-2}$$

$$(\frac{1}{n} + \frac{1}{m})\mathbf{V} \sim W_{n+m-2}((\frac{1}{n} + \frac{1}{m})\Sigma)$$

So $f = n+m-2$ and $a = (\frac{1}{n} + \frac{1}{m})$.
Thus for two-sample T^2 testing $\mathrm{H}_0 : \mu_x - \mu_y = 0$, we have

$$T^2 = (\overline{\mathbf{x}} - \overline{\mathbf{y}})'[(\frac{1}{n} + \frac{1}{m})\mathbf{S}_p]^{-1}(\overline{\mathbf{x}} - \overline{\mathbf{y}})$$

and

$$T^2 \sim \frac{(n+m-2)p}{n+m-p-1}F_{p,n+m-p-1}$$

For large samples,

$$T^2 \sim \chi^2_p$$

Illustrate by comparing first 15 observations to last 15 observations in wood stiffness data.

```
Cmd> X1 <- X[run(15),]

Cmd> X2 <- X[run(16,30),]

Cmd> xbar1 <- tabs(X1,mean:T)

Cmd> xbar2 <- tabs(X2,mean:T)

Cmd> S1 <- tabs(X1,covar:T)

Cmd> S2 <- tabs(X2,covar:T)

Cmd> Sp <- ( (15-1)*S1 + (15-1)*S2)/\
(15+15-2)

Cmd> T2 <- (xbar1-xbar2)'%*%\
solve( (1/15 + 1/15)*Sp) %*% (xbar1-xbar2)

Cmd> T2
(1,1)       4.0808

Cmd> T2/4/(15+15-2)*(15+15-4-1)
(1,1)      0.91089

Cmd> 1-cumF(.91,4,25)
(1)      0.47333
```

In an analogous way, a $1-\alpha$ confidence region for $\mu = \mu_x - \mu_y$ is

$$\left\{\mu : (\mathbf{\bar{x}} - \mathbf{\bar{y}} - \mu)' \left((\frac{1}{n} + \frac{1}{m})\mathbf{S}_p \right)^{-1} (\mathbf{\bar{x}} - \mathbf{\bar{y}} - \mu) \le \frac{(n+m-2)p}{(n+m-p-1)} F_{\alpha,\, p,\, n+m-p-1} \right\}$$

Just as in univariate statistics, assuming equal variances is a strong assumption, and using pooled procedures when variances are unequal gives poor results.
Unpooled variance estimate:

$$\mathbf{V} = \frac{S_x}{n} + \frac{S_y}{m}$$

Under H_0 and for large n and m:

$$T^2 = (\mathbf{\bar{x}} - \mathbf{\bar{y}})'\mathbf{V}^{-1}(\mathbf{\bar{x}} - \mathbf{\bar{y}}) \sim \chi^2_p$$

Likelihood Ratio Tests are a general method in statistics.
Let L be the likelihood as a function of unknown parameters.
Let L_0 be the maximum value of the likelihood when we restrict our parameters to meet the null hypothesis.

Let L_1 be the maximum value of the likelihood over all possibilities.

$$\Lambda = \frac{L_0}{L_1} < 1$$

Λ should be pretty close to 1 if the null is true, but could be arbitrarily small if the null is false. Reject H_0 for small Λ.
For large samples and when H_0 is true

$$-2 \ln \Lambda \sim \chi^2_q$$

where q is the difference in the number of free parameters under the null and alternative hypotheses.
For the T^2 situation, let

$$\hat{\Sigma}_0 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_0)(x_i - \mu_0)'$$

and let

$$\hat{\Sigma}_1 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{\mathbf{x}})(x_i - \overline{\mathbf{x}})'$$

be the maximum likelihood estimates of the variance under H_0 and H_1.
Then

$$L_0 = \frac{e^{-np/2}}{(2\pi)^{n/2} |\hat{\Sigma}_0|^{n/2}}$$

and

$$L_1 = \frac{e^{-np/2}}{(2\pi)^{n/2} |\hat{\Sigma}_1|^{n/2}}$$

and

$$\Lambda = \left(\frac{|\hat{\Sigma}_1|}{|\hat{\Sigma}_0|} \right)^{n/2}$$

Some tedious algebra will show that

$$\frac{|\hat{\Sigma}_1|}{|\hat{\Sigma}_0|} = \frac{1}{1 + \frac{T^2}{n-1}}$$

so that

$$-2 \ln \Lambda = T^2 + O(n^{-1})$$

This is asymptotically χ^2_p, because the alternative includes p additional mean parameters. (But we'd already figured that out another way.)
Where did the $p - 1$ degrees of freedom go in T^2?
Let $w \sim \mathrm{N}_p(0, a\Sigma)$ under H_0.
Let $\mathbf{V} \sim W_f(a\Sigma)$ independent of w.
Find $\mathbf{D}$ such that $\mathbf{D}a\Sigma\mathbf{D} = \mathbf{I}_p$.
Then $w^\star = \mathbf{D}w \sim \mathrm{N}_p(0, \mathbf{I}_p)$ and $\mathbf{V}^\star = \mathbf{D}\mathbf{V}\mathbf{D}' \sim W_f(\mathbf{I}_p)$ (still independent).

$$T^2 = w'\mathbf{V}^{-1}w = w^{\star\prime}\mathbf{V}^{\star-1}w^\star$$

so we can work with the new variables.
Let $\mathbf{Q}_{w^\star}$ be an orthogonal matrix that depends only on $w^\star$. (Drop the $w^\star$ subscript for ease of notation.)

Conditional on $\mathbf{Q}$, $\mathbf{QV}^\star\mathbf{Q}' \sim W_f(\mathbf{QQ}') = W_f(\mathbf{I}_p)$.
Because conditional distribution of $\mathbf{QV}^\star\mathbf{Q}'$ doesn't depend on $\mathbf{Q}$, the unconditional distribution equals the conditional and

$$\mathbf{QV}^\star\mathbf{Q}' \sim W_f(\mathbf{I}_p)$$

$$\begin{aligned} T^2 &= w^{\star\prime}\mathbf{V}^{\star-1}w^\star \\ &= w^{\star\prime}\mathbf{Q}'\mathbf{QV}^{\star-1}\mathbf{Q}'\mathbf{Q}w^\star \\ &= y'\mathbf{B}^{-1}y \end{aligned}$$

where $y = \mathbf{Q}w$, $\mathbf{B} = \mathbf{QV}^\star\mathbf{Q}'$, and y and $\mathbf{B}$ are independent.
Choose the first row of $\mathbf{Q}$ to be $w^{\star\prime}/||w^\star||$; fill in remaining rows in any orthonormal way
Then

$$\mathbf{Q}w^\star = \begin{bmatrix} ||w^\star|| \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

and

$$T^2 = y'\mathbf{B}^{-1}y = ||w^\star||^2\mathbf{B}^{11}$$

where $\mathbf{B}^{11}$ is the 1,1 element of $\mathbf{B}^{-1}$.
$||w^\star||^2 \sim \chi^2_p$
What is the distribution of $\mathbf{B}^{11}$ when $\mathbf{B} \sim W_f(\mathbf{I}_p)$?

$$1/\mathbf{B}^{11} = \mathbf{B}_{11} - \mathbf{B}_{12}\mathbf{B}_{22}^{-1}\mathbf{B}_{21}$$

where

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}$$

and $\mathbf{B}_{11}$ is 1×1, $\mathbf{B}_{12}$ is $1 \times (p-1)$, $\mathbf{B}_{21}$ is $(p-1) \times 1$, and $\mathbf{B}_{21}$ is $(p-1) \times (p-1)$.

$$T^2 = ||w^\star||^2\mathbf{B}^{11} = \chi^2_p\mathbf{B}^{11} = \chi^2_p/[\mathbf{B}_{11} - \mathbf{B}_{12}\mathbf{B}_{22}^{-1}\mathbf{B}_{21}]$$

If $\mathbf{B} \sim W_f(\mathbf{I}_p)$, then

$$\mathbf{B}_{11} - \mathbf{B}_{12}\mathbf{B}_{22}^{-1}\mathbf{B}_{21} \sim \chi^2_{f-(p-1)}$$

Thus we get a ratio of chisquared distributions for T^2, and an F distribution after suitable rescaling via degrees of freedom.
The distributional result can be modified for $W_f(\Sigma)$, and modified for a submatrix bigger than 1×1 (we'll get a Wishart). But you always lose a degree of freedom for every variable left out of the submatrix.
For you folks in 8401, try to prove the following:
Theorm. Suppose that $y_1, y_2, \ldots, y_m$ are independent with $y_i \sim \mathrm{N}_p(\Gamma w_i, \Sigma)$, where Γ is a fixed matrix and w_i is some r-vector. Let $\mathbf{H} = \sum_{i=1}^m w_i w_i'$ and assume that $\mathbf{H}$ is nonsingular. Let $\mathbf{G} = \sum_{i=1}^m y_i w_i'\mathbf{H}^{-1}$. Then

$$\sum_{i=1}^m y_i y_i' - \mathbf{GHG}' \sim W_{m-r}(\Sigma)$$

independent of $\mathbf{B}$.

Hint: Let $\mathbf{W}$ be the $r \times m$ matrix with columns w_i, let $\mathbf{F}$ be square such that $\mathbf{FHF}' = \mathbf{I}$, let $\mathbf{E}_2 = \mathbf{FW}$. Complete $\mathbf{E}_2$ to a full $m \times m$ orthogonal matrix $\mathbf{E}$

$$\mathbf{E} = \begin{bmatrix} \mathbf{E}_1 \\ \mathbf{E}_2 \end{bmatrix}$$

Let $u = y\mathbf{E}'$, and work with the u vector.

Corollary. Let $\mathbf{P} = (n-1)\mathbf{S}$ be the matrix of sums of squares and cross products from an iid sample y_i from $\mathrm{N}_p(\mu, \Sigma)$. Partition $\mathbf{P}$ into its first q rows and columns and the remaining $p-q$ rows and columns. Define

$$\mathbf{P}_{11\bullet 2} = \mathbf{P}_{11} - \mathbf{P}_{12}\mathbf{P}_{22}^{-1}\mathbf{P}_{21}$$

and

$$\Sigma_{11\bullet 2} = \Sigma_{11} - \Sigma_{12}\Sigma P_{22}^{-1}\Sigma_{21}$$

Then

$$\mathbf{P}_{11\bullet 2} \sim W_{n-1-(p-q)}(\Sigma_{11\bullet 2})$$

Hint: Find the conditional distribution of the first q elements of y_i conditional on the last $p-q$. Then use the preceding theorem.